PHD

Analysis of power functions of multiple comparisons tests

Liu, Wei

Award date:
1990

Awarding institution:
University of Bath

Analysis of Power Functions
of
Multiple Comparisons Tests

Submitted by
Wei Liu
for the degree of PhD
of the
University of Bath
1990

To my father and mother

献给我的父母

Analysis of Power Functions of Multiple Comparisons Tests

Submitted by
Wei Liu
for the degree of PhD
of the
University of Bath
1990

Copyright

Wei Liu

W. Liu

Analysis of Power Functions of Multiple Comparisons Tests

Abstract

The power functions of the Studentized range test and the F-test for comparing the means of normal populations in the one-way fixed effects analysis of variance model is investigated. The least favourable configurations of population means under certain range restrictions of the population means are given. These results are important in the calculation of the sample sizes required to guarantee power level under the restrictions on the range of the population means. The optimal test procedures which maximise the least favourable power under the range restrictions of the population means are found for the known variance case. Then the relations between these optimal tests and the range type tests, sum of square tests are pointed out. The problem of comparing several treatment means with a control mean is also considered. Finally, these results are applied to the binary data, by using the arcsin-root transformation and the asymptotic distribution, for both comparison between k Bernoulli responses and comparison between k Bernoulli responses and a control Bernoulli response.

Acknowledgements

I am very grateful to my supervisor, Dr. A.J. Hayter, for his guidance and help during the course of this work. I would also like to thank all the staff and postgraduates of the School of Mathematical Sciences, University of Bath. I acknowledge the financial support of the British Council and the Chinese Government.

Table of Contents

Chapter 1

Introduction

1.1 Simultaneous test procedures

One problem commonly encountered in empirical research is that of the comparison of k treatment means, with $k \geq 3$. A possible approach to the detection of any differences might be to perform C_2^k separate pairwise tests, each at significance level α. However, although each individual test in this procedure has size α, the significance level for the overall homogeneity hypothesis will be greater. To cope with this difficulty, various simultaneous test procedures have been developed for different situations, all of which guarantee the overall significance level, for example, Scheffe's F-test, Tukey's Studentised range test, and Dunnett's many-one tests. Miller's celebrated book "Simultaneous statistical inference" and Hochberg & Tamhane's "Multiple comparison procedures" have both made a good summary of the procedures available and discuss the latest developments in this field.

As with standard hypothesis tests, simultaneous test procedures are constrained by error rate considerations, namely, the error rate under the null hypothesis---the significance level, and the error rate under the alternative hypothesis--- one minus the power. So far, the research in simultaneous test procedures has usually concentrated on the former, partly for the reason that it can be used to derive simultaneous confidence intervals. As far as the power is concerned, any study is usually much more difficult; this is because the distribution under the alternative hypothesis tends to be more complicated than that under the null hypothesis. For the power function of the F-test, because under the alternative hypothesis the test statistic has a non-central F distribution which depends on the population means simply through the non-centrality parameter, a large amount of work has been available in literature, for instance, Pearson & Hartley (1951), Fox (1956), Tiku (1967), Dasgupta (1968) and Kastenbaum, Hoel & Bowman (1970). However for the power of the other simultaneous tests, Miller once said; "it seems fruitless to pursue this in detail, since the simplest problem of determining the probability of rejecting the overall null hypothesis under an alternative has been solved only for the F and t statistics. This last sentence essentially summarizes our knowledge on power for multiple comparisons techniques. The power functions of F and t statistics are pretty well known today. Those for the Studentised range, the Studentised maximum modulus, and the many-one t statistics are unknown. This is due to the obstacle that the power functions of these latter statistics are not functions of single noncentrality parameters which are simple functions of the theoretical means, as in the case of the F statistic" (see Miller (1981), p102). More than twenty years have passed since Miller said these words, but the progress has been, and still is, very slow. To the author's knowledge, the main work available for the power function of the Studentised range test, one of the most important simultaneous test procedures, is David, Lachenbruch & Brandis (1972) in which they have performed some computer calculations of the power of the Studentised range test in certain situations. In this thesis, the power properties of some important simultaneous test procedures such as the Studentised range test are studied, and some new simultaneous test procedures derived. In the following section, some basic notation is introduced.

Suppose the sample $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ has distribution $F\ (\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\zeta}\ )$, where $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in \mathbf{R}^k$, and $(\boldsymbol{\theta}, \boldsymbol{\zeta}) = ((\theta_1, \theta_2, \ldots, \theta_r), (\zeta_1, \zeta_2, \ldots, \zeta_s))$ are unknown parameters with parameter space $\Theta \times Z \subseteq \mathbf{R}^r \times \mathbf{R}^s$. We are interested in the parameters $\boldsymbol{\theta}$, while $\boldsymbol{\zeta}$ are nuisance parameters. Assume Θ_0 is a non-empty subset of Θ. Denote the complement of Θ_0 in Θ by $\Theta_a = \Theta - \Theta_0$. Consider the (simultaneous) hypothesis test problem

$$\textbf{the null hypothesis} \qquad H_0 : \boldsymbol{\theta} \in \Theta_0$$

against (1.1)

$$\textbf{the alternative hypothesis} \qquad H_a : \theta \in \Theta_a .$$

A test procedure is to partition the set $\Omega \subseteq \mathbf{R}^k$ of all possible values of $\mathbf{X}$ (the sample space) into two parts $\mathbf{C}$ and $\Omega - \mathbf{C}$, such that

$$\mathbf{X} \in \mathbf{C} \quad \Rightarrow \quad \text{reject } H_0 ,$$

$$\mathbf{X} \in \Omega - \mathbf{C} \quad \Rightarrow \quad \text{accept } H_0 .$$

So, a test procedure is specified by $\mathbf{C}$, the **critical region** of the test. A **test statistic** is a statistic that is used in defining $\mathbf{C}$.

One can make two types of error in a test procedure :

Type I error : Reject H_0 when H_0 is true.

Type II error : Accept H_0 when H_0 is false.

Ideally, we would like both types of error to have probability 0, however this is impossible except in trivial cases. We therefore have to content ourselves with trying to keep these error probabilities at an acceptably small level. It is customary, to fix the probability of type I error at a preassigned (small) level α, $0<\alpha<1$, and then to try to minimize the probability of type II error. This fixing of α at small level reflects our confidence in the choice of H_0

For a test with critical region $\mathbf{C}$, the value α which satisfies :

$$\sup_{\theta \in \Theta_0} P_\theta \text{ (reject } H_0 \text{)} = \sup_{\theta \in \Theta_0} P_\theta (\mathbf{X} \in \mathbf{C}) = \alpha$$

is called the **significance level** (or **size**) of the test. In order to assess the performance of the test, once the significance level has been fixed, we need to consider the probability of the type II error. If we define :

$$\beta(\theta) = P_\theta(\text{ reject } H_0) = P_\theta(\mathbf{X} \in \mathbf{C}) ,$$

then as a function of θ, $\beta(\theta)$ is called **the power function** of the test. Notice that if $\theta \in \Theta_0$ then $\beta(\theta)$ is the probability of a type I error, and if $\theta \in \Theta_a$ then $\beta(\theta)$ equals 1 - P (type II error). Thus minimization of a type II error probability over different tests is equivalent to the maximization of $\beta(\theta)$ for $\theta \in \Theta_a$ over different tests.

Theoretically, amongst all the tests with significance level α, we want to find the test whose power function $\beta^*(\theta)$ satisfies

$$\beta^*(\theta) \geq \beta(\theta), \text{ for any } \boldsymbol{\theta} \in \Theta_a ,$$

where $\beta(\boldsymbol{\theta})$ is the power function of any test with significance level α. Such test is called the **uniformly most powerful test** (UMPT). However the UMPT does not exist for many hypothesis testing problems, especially for multiple comparison test problems. So, we have to content ourselves with some "reasonable" test procedures for most multiple comparison problems. In the next section, we introduce ideas which will be used to investigate the power properties of the available test procedures, and also the notion of a maximin test procedure.

1.2 Partitioning of the alternative hypothesis

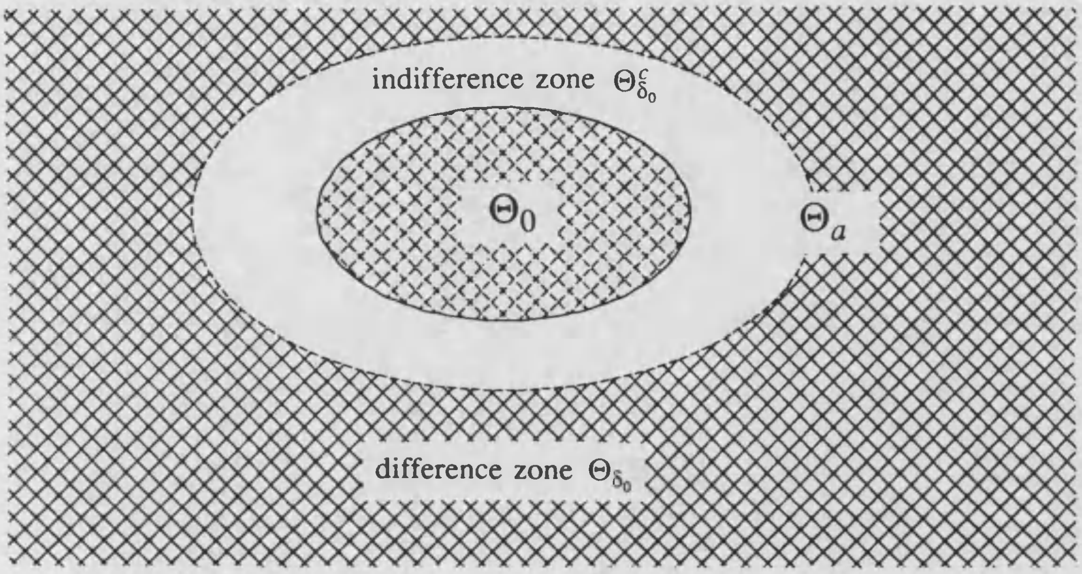
For any reasonable test procedure for the hypothesis test problem (1.1) at significance level α, the power function $\beta(\theta)$ will usually be a continuous function of $\theta \in \Theta$. Thus we have

$$\inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_a} \beta(\theta) \leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta(\theta) = \alpha \text{ (small) } .$$

The magnitude of $\beta(\theta)$, for $\boldsymbol{\theta} \in H_a$, will also depends on the distance between the point θ and the null hypothesis parameter set Θ_0, i.e. the further θ is from Θ_0, the larger is its power value. The value $\beta(\theta)$ could be as small as α when $\boldsymbol{\theta} \in \Theta_a$ approaches Θ_0. The implication of this is that, for those points $\theta \in \Theta_a$ very close to Θ_0, it may be impossible, whatever the sample size, to meet some power requirement $\beta(\theta) \geq \beta_0$ where β_0 is a prefixed positive number in $(0, 1)$ usually greater than α. So a natural modification is to guarantee the power value only for those $\boldsymbol{\theta} \in \Theta_a$ which are sufficiently far from Θ_0; in practical situations there would not be any serious problem if we fail to identify those points which are close to, but not members of, Θ_0. In other words, we can partition the alternative hypothesis parameter set Θ_a into two parts, one made up of those points which lie sufficiently far from Θ_0, the other consisting of the remaining points of Θ_a. Then we guarantee the power value only for those $\boldsymbol{\theta}$ in the first part. To partition Θ_a, we can specify a function $d(\boldsymbol{\theta}, \Theta_0)$ which measures the distance between $\boldsymbol{\theta}$ and Θ_0, and a positive number δ_0. Then partition Θ_a into the parts :

$$\Theta_{\delta_0} \equiv \{ \ \theta : \ d(\boldsymbol{\theta}, \Theta_0) \geq \delta_0 \ \} \quad \text{ and } \quad \Theta^c_{\delta_0} \equiv \Theta_a - \Theta_{\delta_0} ,$$

i.e. Θ_{δ_0} is made of the points whose distance from Θ_0, as defined by $d(\theta,\Theta_0)$, is at least δ_0, and $\Theta_{\delta_0}^c$ is made from the remaining points whose distance from Θ_0 is less than δ_0. The power requirement is then guaranteed only for Θ_{δ_0}. Notice that, in this approach to the power, we are interested only in Θ_{δ_0} instead of the whole alternative parameter set. So we call Θ_{δ_0} the **difference zone**, and $\Theta_{\delta_0}^c$ the **indifference zone** (this is the same notation as is used in Bechhofer's(1954) indifference approach in ranking and selection).



To illustrate the basic idea of the above, it is convenient to take a simple case: Let $X \sim N(\theta, 1/n)$, where X can be regarded as the sample mean and n as the sample size. $\Theta = \mathbf{R}$, $\Theta_0 = \{0\}$, $\Theta_a = \Theta - \Theta_0 = \{ \theta: \theta \neq 0 \}$. So, the problem is to test

$$H_0: \theta = 0 \qquad \text{against} \qquad H_a: \theta \neq 0.$$

One possible α-level test has critical region

$$\mathbf{C} = \{ X: | \sqrt{n}\, X | > \lambda_{\alpha/2} \},$$

where $\lambda_{\alpha/2}$ is the $\alpha/2$ quantile of a standard normal distribution. Then the power function of this test is

$$\begin{aligned}\beta(\theta) &= P_\theta \{ X \in \mathbf{C} \} = P_\theta \{ | \sqrt{n}\, X | > \lambda_{\alpha/2} \} \\ &= P \{ | N(\sqrt{n}\,\theta, 1) | > \lambda_{\alpha/2} \} \\ &= 1 - \int_{-\sqrt{n}\theta-\lambda_{\alpha/2}}^{-\sqrt{n}\theta+\lambda_{\alpha/2}} \varphi(x)\, dx,\end{aligned}$$

where $\varphi(x)$ is the p.d.f. of a standard normal random variable. It is easy to see that

$$\lim_{\theta \to \theta_0 = 0} \beta(\theta) = \alpha,$$

whatever the sample size n. In other words, for those θ close to $\theta_0=0$, the probability of correct classification and thus of rejection of H_0 cannot be guaranteed to be any greater than the significance level α, no matter how large the sample size. Fortunately, it is not of great consequence in empirical research if we cannot identify those points whose distance from zero is less than some small positive number δ_0 (whose magnitude will depends upon the problem under consideration); we guarantee the power value only for those points in $\Theta_{\delta_0} = \{ \theta : |\theta| \geq \delta_0 \}$. Since

$$\inf_{\theta \in \Theta_{\delta_0}} \beta(\theta) = \beta(\delta_0) = 1 - \int_{-\sqrt{n}\,\delta_0 - \lambda_{\alpha/2}}^{-\sqrt{n}\,\delta_0 + \lambda_{\alpha/2}} \varphi(x)\, dx ,$$

and as n increases to infinity, the integral

$$\int_{-\sqrt{n}\,\delta_0 - \lambda_{\alpha/2}}^{-\sqrt{n}\,\delta_0 + \lambda_{\alpha/2}} \varphi(x)\, dx$$

decreases to zero, by a suitable choice of n, we can have

$$\inf_{\theta \in \Theta_{\delta_0}} \beta(\theta) \geq \beta_0 ,$$

where β_0 is an arbitrary prefixed number in $(\alpha, 1)$. So we can uniformly guarantee the required power value for those points in Θ_{δ_0} by picking a sufficiently large sample size.

Through this simple example we can see that, for the proposed null hypothesis test problem (1.1), we should carry out the following three main steps.

(i) First we need a reasonable test procedure, although we will not usually know whether the employed procedure is optimal. Many procedures suitable for different situations have already been developed, see for examples, Miller and Hochberg & Tamhane's books. In addition we have developed some new simultaneous test procedures, and studied their power properties following the above ideas.

(ii) Secondly the alternative hypothesis parameter set Θ_a must be partitioned into two, the difference and the indifference zone, in order to be able to guarantee the power of the employed test procedure just for the parameters in the difference zone. For this, we need a function $d(\theta, \Theta_0)$ which measures the distance between θ and Θ_0, thus partitioning Θ_a into

the difference zone $\quad \Theta_{\delta_0} \equiv \{ \theta : d(\theta, \Theta_0) \geq \delta_0 \ , \theta \in \Theta_a \},$

and

the indifference zone $\quad \Theta_{\delta_0}^c \equiv \{ \theta : d(\theta, \Theta_0) < \delta_0, \theta \in \Theta_a \},$

where δ_0 is a prefixed positive number. The choice of this measure function depends upon the problem under consideration, and for a single problem there may be several possible suitable choices. For example, in the case of testing the null hypothesis H_0, the equality of all the k treatment means $\theta_1 = \theta_2 = \ldots = \theta_k$, against the alternative hypothesis "not H_0", $d(\theta, \Theta_0)$ could take any of the following forms :

$$d_1(\theta, \Theta_0) = \max_{1 \leq i,j \leq k} | \theta_i - \theta_j | ,$$

$$d_2(\theta, \Theta_0) = \max_{1 \leq i \leq k} | \theta_i - \bar{\theta} |, \quad \text{where} \quad \bar{\theta} = \sum_{i=1}^{k} \theta_i / k ,$$

$$d_3(\theta, \Theta_0) = \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2 .$$

Classically, $d_3(\theta, \Theta_0) = \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2$ is used in the literature because of its relationship with the F-test (see chapter 2). It would seem logical that the definition of $d(\theta, \Theta_0)$ should not depend on the choice of test procedure, but rather on the structure of Θ_0 and Θ_a and the practical requirements of the experimenter, i.e what is the most suitable choice of difference zone for the problem under consideration.

(iii) Thirdly, to guarantee the power value for those θ in the difference zone Θ_{δ_0}, i.e. so that

$$\inf_{\theta \in \Theta_{\delta_0}} \beta(\theta) \geq \beta_0 , \tag{1.2}$$

where β_0 is a preassigned positive number in $(\alpha, 1)$. We define the $\theta^* \in \Theta_{\delta_0}$ which satisfies

$$\inf_{\theta \in \Theta_{\delta_0}} \beta(\theta) = \beta(\theta^*), \tag{1.3}$$

as the **least favourable configuration** for the difference zone Θ_{δ_0}, and the respective power value as the **least favourable power**. The least favourable configuration usually depends not only on the choice of the difference zone but also on the test procedure employed, as does the least favourable power. For the given test procedure and given difference zone, the least favourable configuration may not be unique, however by its definition the least favourable power can only take one value. For those problems for which we can find an explicit form for at least one of the least favourable configurations, we can guarantee the power level for the difference zone simply by fixing the least favourable power to be no less than β_0. When we cannot find the explicit form for any of the least favourable configurations, it is usually possible to shrink Θ_{δ_0} to a subset of itself, $\Theta^{s}_{\delta_0}$ say, which is known to contain at least one of the least favourable configurations. Thus we can guarantee the inequality in (1.2) simply by guaranteeing that :

$$\inf_{\theta \in \Theta^{s}_{\delta_0}} \beta(\theta) \geq \beta_0 .$$

This can sometimes be done using a numerical search technique.

Steps (i),(ii) and (iii) above constitute the partition approach to the problem (1.1). However, notice that for the same problem and the same difference zone, different test procedures may have different least favourable powers. If in a properly defined family of test procedures all with size α, there is a test procedure whose least favourable power is no less than the least favourable power of any other test procedure in this family, then we call this one the **maximin test procedure** (of that test family). Obviously, the maximin test procedure is the optimal test for our approach in the sense that it guarantees the maximum value of the least favourable power in the considered test family. We have found the maximin tests for some cases, but in most situations, the maximin test procedure is difficult to identify.

1.3 On the chapters to follow

Chapter 2 deals with the two best known simultaneous test procedures, i.e. Scheffe's F-test and Tukey's Studentised range test, both for comparing k

treatment means : $\mu_1, \mu_2, \ldots, \mu_k$ in a one-way fixed effects analysis of variance model. We first derive least favourable configurations of the F-test and the Studentised range test under the distance measure functions $\max_{1 \le i,j \le k} | \mu_i - \mu_j |$ and $\max_{1 \le i \le k} | \mu_i - \bar{\mu} |$, where $\bar{\mu} = \sum_{i=1}^{k} \mu_i / k$. Then we develop the method to compute the least favourable powers, in order to calculate the sample size necessary to guarantee required least favourable powers for the difference zones

$$\{ \mu : \max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge \delta_0 \sigma \} \quad \text{and} \quad \{ \mu : \max_{1 \le i \le k} |\mu_i - \bar{\mu}| \ge \delta_0 \sigma \}.$$

Finally, in section 2.4, we consider some two-stage sample procedures in order to guarantee the power requirement for the difference zones

$$\{ \mu : \max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge \delta_0 \} \quad \text{and} \quad \{ \mu : \max_{1 \le i \le k} |\mu_i - \bar{\mu}| \ge \delta_0 \}.$$

In chapter 3, we first derive the least favourable configurations for a general family of distributions and a general family of test procedures for comparing k treatment means. In the particular case of comparing the means of k independent normal distributions with equal known variances, the maximin test procedure is found. The relation between this maximin test procedure and the χ^2-test (the special case of the F-test when the variance is known), and the range test (the special case of the Studentised range test when the variance is known) is pointed out.

Chapter 4 is devoted to testing the equality of several proportions. By using the large sample property of the arcsin-root transformation and the results of chapter 3, a simultaneous test procedure is developed, and a method for computing the exact size and exact power at the asymptotic least favourable configuration is derived.

In chapter 5, we consider the problem of comparing k treatment means with a control mean. First, we investigate the power properties of Dunnett's many-one test procedures. By using that property and the arcsin-root transformation again, we develop a simultaneous test procedure for the comparison of k treatment proportions with a control proportion.

In chapter 6, we give a summary, and point out some possible directions for research in the future.

Chapter 2

The power functions of the Studentised range test and the F-test

---

2.1 The Studentised range test and the F-test

Consider the usual balanced one-way fixed effects analysis of variance model

$$X_{ij} = \mu_i + \varepsilon_{ij} , \qquad 1 \le i \le k, \ 1 \le j \le n , \tag{2.1}$$

where the μ_i, $1 \le i \le k$, are the k unknown population means, and the ε_{ij} are independently, identically distributed as $N(0,\sigma^2)$ random variables for some unknown error variance σ^2. Define $\mu = (\mu_1, \mu_2, \cdots, \mu_k)$ to be the vector of the k population means. The sample means of the k populations are given by :

$$\bar{X}_i = \frac{\sum_{j=1}^{n} X_{ij}}{n} \sim N (\mu_i, \sigma^2/n), \qquad 1 \le i \le k .$$

It is assumed that an estimate S^2 of σ^2 is available with the following distribution : $S^2 \sim \sigma^2 \dfrac{\chi_\nu^2}{\nu}$ for some degrees of freedom ν, independent of the k sample means $\bar{X}_i$, $1 \le i \le k$. Usually, the analysis of variance mean square

$$S^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2}{k(n-1)}$$

will be used with $\nu = k(n-1)$. We are interested in making inferences on the population means μ_i, $1 \le i \le k$, and in particular in testing the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

against the general alternative hypothesis of "not H_0". We shall consider the Studentised range test and the F-test for this null hypothesis; details of these and alternative procedures may be found in Hochberg and Tamhane (1987).

The Studentised range test of this null hypothesis operates as follows. Define

$$Q_{k,\nu}(\mu) = \frac{(\max\ |\bar{X}_i - \bar{X}_j| : 1 \le i,j \le k)\sqrt{n}}{S}$$
$$= \frac{(\bar{\mathrm{X}}_{\max} - \bar{\mathrm{X}}_{\min})\sqrt{n}}{S},$$

where $\bar{\mathrm{X}}_{\max}$ ($\bar{\mathrm{X}}_{\min}$) is the maximum (minimum) of the k sample means, $\bar{X}_i$, $1 \le i \le k$. A size α test is obtained by rejecting the null hypothesis H_0 iff

$$Q_{k,\nu}(\mu) \ge q_{k,\nu}^{\alpha},$$

where $q_{k,\nu}^{\alpha}$ is the upper α-point of the Studentised range distribution with parameters k and ν. Such a test has size exactly equal to the nominal value α under the null hypothesis H_0. The power function of the Studentised range test is

$$\beta_S(\mu/\sigma, \nu, \alpha) = P(Q_{k,\nu}(\mu) \ge q_{k,\nu}^{\alpha}).$$

Notice that we denote this power function by $\beta_S(\mu/\sigma, \nu, \alpha)$ because it depends on μ and σ only through μ/σ (see section 2.2). It is of great interest to discover exactly how this power function depends on μ, the vector of the k population means. However, this dependence is not simple and has hitherto defied investigation. David, Lachenbruch & Brandis (1972) have made computer calculations of this power function under certain configurations of the population means vector μ.

An alternative method of testing the null hypothesis H_0 is to use the F-test which operates as follows. Define

$$L_{k,\nu}(\mu) = \frac{n \sum_{i=1}^{k} (\bar{X}_i - \bar{\bar{X}})^2}{(k-1)\, S^2},$$

where

$$\bar{\bar{X}} = \frac{\sum_{i=1}^{k} \bar{X}_i}{k}.$$

Then the F-test of size α is to reject the null hypothesis H_0 iff

$$L_{k,\nu}(\mu) \quad \geq \quad F^{\alpha}_{k-1,\nu} \ ,$$

where $F^{\alpha}_{k-1,\nu}$ is the upper α-point of the F-distribution with parameters $k-1$ and ν. Again this test has size exactly equal to α under the null hypothesis H_0. The power function of the F-test is

$$\begin{aligned} \beta_F(\mu/\sigma,\ \nu,\alpha) &= P\ (\ L_{k,\nu}(\mu) \geq F^{\alpha}_{k-1,\nu}\ ) \\ &= P\ (\ F_{k-1,\nu}\ (\delta) \geq F^{\alpha}_{k-1,\nu}\ )\ , \qquad (2.2) \end{aligned}$$

where $F_{k-1,\nu}\ (\delta)$ denotes a non-central F random variable with non-centrality parameter δ and degrees of freedom $k-1$ and ν, and

$$\delta^2 \quad = \quad \frac{n \sum_{i=1}^{k} (\mu_i\ -\ \bar{\mu})^2}{\sigma^2}$$

with

$$\bar{\mu} \quad = \quad \frac{\sum_{i=1}^{k} \mu_i}{k} .$$

Thus the power function of the F-test depends on the vector of population means μ only through the quantity

$$b_F(\mu) \quad = \quad \sum_{i=1}^{k} (\mu_i\ -\ \bar{\mu})^2\ . \qquad (2.3)$$

This simple dependence makes the investigation of the properties of the power function easy and has allowed the calculation of tables of the power function (see, for example, Fox (1956) and Tiku (1967)).

In next section we obtain some theoretical results about the power functions of the Studentised range test and the F-test, in particular the explicit forms of their common least favourable configurations for the two difference zones

$$\{\ \mu\ :\ \max_{1\leq i\leq k} |\mu_i - \bar{\mu}| \geq b_0\sigma\ \} \quad \text{and} \quad \{\ \mu\ :\ \max_{1\leq i,j\leq k} |\mu_i - \mu_j| \geq b_0\sigma\ \}.$$

In section 2.3 we discuss the problem of designing experiments to meet power requirements for these two difference zones. The method to calculate the least favourable powers of the Studentised range test and F-test has been presented. Some comparisons between the least favourable power of the Studentised range test and the least favourable power of the F-test show that neither power function dominates the other, in the sense that for some sets of population means μ, the Studentised range test is more powerful than the F-test, while for other sets of population means, the F-test is more powerful than the Studentised range test. In section 2.4 we consider the the problem of guaranteeing the power level for the difference zones

$$\{ \mu : \max_{1\le i\le k} |\mu_i - \bar{\mu}| \ge b_0 \} \text{ and } \{ \mu : \max_{1\le i,j\le k} |\mu_i - \mu_j| \ge b_0 \}$$

by using two-stage sample procedures.

2.2 The power functions of the Studentised range test and the F-test

In this section we present some theoretical results concerning the power functions of the Studentised range test and the F-test.

Notice that the power function of the Studentised range test can be written as

$$\begin{aligned}
\beta_S(\mu/\sigma, \nu, \alpha) &= P(\, Q_{k,\nu}(\mu) \ge q^{\alpha}_{k,\nu}\,) \\
&= P\left[\frac{(\bar{X}_{\max} - \bar{X}_{\min})\sqrt{n}}{S} \ge q^{\alpha}_{k,\mu} \right] \\
&= \int_{s=0}^{\infty} P\{ \frac{(\bar{X}_{\max} - \bar{X}_{\min})\sqrt{n}}{\sigma\sqrt{s}} \ge q^{\alpha}_{k,\nu} \} \, f_{\chi^2_\nu/\nu}(s)\, ds \\
&= 1 - \int_{s=0}^{\infty} P\{ \frac{(\bar{X}_{\max} - \bar{X}_{\min})\sqrt{n}}{\sigma\sqrt{s}} \le q^{\alpha}_{k,\nu} \} \, f_{\chi^2_\nu/\nu}(s)\, ds \\
&= 1 - \int_{s=0}^{\infty} W(\sqrt{n}\,\mu/\sigma, \sqrt{s}\, q^{\alpha}_{k,\nu})\, f_{\chi^2_\nu/\nu}(s)\, ds\,, \qquad (2.4)
\end{aligned}$$

where $f_{\chi^2_\nu/\nu}(s)$ is the probability density function of a χ^2_ν/ν random variable, and the function $W(\theta, c)$ for $\theta \in \mathbf{R}^k$ and $c \in \mathbf{R}$ is defined as

$$W(\theta, c) = P\left\{ |Y_i - Y_j| \le c \; ; \; 1 \le i,j \le k \right\}, \qquad (2.5)$$

and where the Y_i, $1 \le i \le k$, are independent random variables with

$$Y_i \sim N(\theta_i, 1)\,, \qquad 1 \le i \le k\,. \qquad (2.6)$$

Thus, in order to determine how the power function $\beta_S(\mu/\sigma, \nu, \alpha)$ depends on μ, it is sufficient to determine how the function $W(\theta, c)$ depends on θ, for any positive real number c. For instance, it is clear that for any $c \in \mathbf{R}$,

$$W(\boldsymbol{\theta},c) = W(-\boldsymbol{\theta},c) \ , \tag{2.7a}$$

$$W(\boldsymbol{\theta}+\lambda\mathbf{1},c) = W(\boldsymbol{\theta},c) \ , \tag{2.7b}$$

$$W(\pi(\boldsymbol{\theta}),c) = W(\boldsymbol{\theta},c) \ , \tag{2.7c}$$

where $\mathbf{1} = (1, \cdots, 1) \in \mathbf{R}^k$, and the operator π permutes coordinates. Therefore it follows from equation (2.4) that similar properties hold for the power function of the Studentised range test, i.e.

$$\beta_S(\boldsymbol{\mu}/\sigma, \nu,\alpha) = \beta_S(-\boldsymbol{\mu}/\sigma, \nu,\alpha) \ , \tag{2.8a}$$

$$\beta_S((\boldsymbol{\mu}+\lambda\mathbf{1})/\sigma, \nu,\alpha) = \beta_S(\boldsymbol{\mu}/\sigma, \nu,\alpha) \ , \tag{2.8b}$$

$$\beta_S(\pi(\boldsymbol{\mu})/\sigma, \nu,\alpha) = \beta_S(\boldsymbol{\mu}/\sigma, \nu,\alpha) \ . \tag{2.8c}$$

So, reflection of $\boldsymbol{\mu}$ about the origin, or shifting or exchanging the coordinates of $\boldsymbol{\mu}$ will not change the power of the Studentised range test. In fact these three properties are also true for the power function of the F-test, as can easily be seen from equation (2.3). Thus for the Studentised range test and the F-test, these three operations (or combinations thereof) will take one least favourable configuration to another, provided that the transformed mean vector still belongs to the difference zone under consideration.

We now investigate the function $W(\boldsymbol{\theta},c)$, which can be written as

$$W(\boldsymbol{\theta},c) = \int_{\omega} g(\mathbf{y}-\boldsymbol{\theta}) \, d\mathbf{y} \ , \tag{2.9}$$

where

$$\omega = \{ \ \mathbf{y} = (y_1, \cdots, y_k) : \ |y_i - y_j| \le c \ ; \ 1 \le i,j \le k \ \} \subset \mathbf{R}^k \ , \tag{2.10}$$

and

$$g(\mathbf{x}) = \prod_{i=1}^{k} \varphi(x_i) \ , \tag{2.11}$$

where $\varphi(.)$ is the probability density function of a $N(0,1)$ random variable.

Theorem 2.1 below enables us to compare the power value of the Studentised range test for cases in which one set of means $\boldsymbol{\mu}^*$ majorises another set $\boldsymbol{\mu}$. The definition of majorisation is given in the appendix.

Theorem 2.1 Suppose $\boldsymbol{\mu}, \boldsymbol{\mu}^* \in \mathbf{R}^k$, and suppose that the vector $\boldsymbol{\mu}^*$

majorises the vector μ, i.e. $\mu^* \gg \mu$. Then, for all $\sigma > 0$, $\nu \in \mathbf{N}$ and $0 \le \alpha \le 1$, we have

$$\beta_S(\mu^*/\sigma, \nu,\alpha) \ \ge\ \beta_S(\mu/\sigma, \nu,\alpha).$$

Proof of Theorem 2.1

It follows from equation (2.4) that in order to prove Theorem 2.1, it is sufficient to show that, for all $c \in \mathbf{R}$,

$$\mu* \gg \mu \quad \Rightarrow \quad W(\mu*,c) \le W(\mu,c). \tag{2.12}$$

Now, the function $g(\mathbf{x})$ defined in equation (2.11) is well known to be a Schur-concave function (see definition in the appendix), and the set ω defined in equation (2.10) is easily seen to be permutation invariant and convex. Thus, equation (2.12) follows from equation (2.9) and Theorem A1 and Note A1 contained in the appendix, and this completes the proof of Theorem 2.1 #

Notice that majorisation applies only to a subset of all possible μ, so Theorem 2.1 cannot be used to compare power for any two mean vectors. However, the following corollaries to Theorem 2.1 can be used to compare the respective power values of two sets of means in some cases other than majorisation.

Corollary 2.1 Let $\mu = (\mu_1, \ldots ,\mu_k) \in \mathbf{R}^k$ with $\mu_1 \le \mu_2 \le \ldots \le \mu_k$, and let $\delta = (\delta_1, \ldots ,\delta_k)$ with $\delta_1 \le \delta_2 \le \ldots \le \delta_k$. Then, for all $\sigma > 0$, $\nu \in \mathbf{N}$ and $0 \le \alpha \le 1$, we have

$$\beta_S((\mu+\delta)/\sigma, \nu,\alpha) \ \ge\ \beta_S(\mu/\sigma, \nu,\alpha).$$

Proof of Corollary 2.1

Define $\mu^* \ = \ (\mu_1+\delta_1-\bar{\delta} , \ldots , \mu_k+\delta_k-\bar{\delta})$ where $\bar{\delta} \ = \ \frac{1}{k} \sum_{i=1}^{k} \delta_i$. Then, from equation (2.8b) with $\lambda=-\bar{\delta}$, we have that

$$\beta_S((\mu+\delta)/\sigma, \nu,\alpha) \ = \ \beta_S(\mu^*/\sigma, \nu,\alpha).$$

Also, we have $\mu^* \gg \mu$, and so the corollary follows from Theorem 2.1 #

Corollary 2.2 Let $\mu = (\mu_1, \ldots ,\mu_k) \in \mathbf{R}^k$, and let the real number $\lambda \ge 1$. Then, we have

$$\beta_S(\lambda\, \mu/\sigma, \nu,\alpha) \ \ge\ \beta_S(\mu/\sigma, \nu,\alpha).$$

Proof of Corollary 2.2

Follows immediately from Corollary 2.1 with $\delta_i = (\lambda-1)\mu_i$ and using permutation invariance (2.8c) #

Now we are going to present the main results of this section which give the least favourable configurations of the power functions of the Studentised range test and the F-test for the difference zones

$$\Theta^1_{b_0} \equiv \{ \mu : b_1(\mu) \equiv \max_{1\le i\le k} |\mu_i - \bar{\mu}| \ge b_0 \sigma \} \tag{2.13}$$

where $\bar{\mu} = \sum_{i=1}^{k} \mu_i / k$, and

$$\Theta^2_{b_0} \equiv \{ \mu : b_2(\mu) \equiv \max_{1\le i,j\le k} |\mu_i - \mu_j| \ge b_0 \sigma \} , \tag{2.14}$$

with $b_0 \ge 0$. As we shall see later, the results in this section are special cases of Theorem 3.1 and Theorem 3.2 of chapter 3; the proofs in this section, which use the log-concavity of $W(\theta,c)$ as a function of θ, are different from those in chapter 3 however. The definition of log-concavity is given in the appendix. In order to prove Theorem 2.2 and Theorem 2.4, we need the following lemma, 2.1.

Lemma 2.1

(i) $W(\theta,c)$ is log-concave, $i.e.$ for $0\le\alpha\le1$,

$$W(\alpha\theta+(1-\alpha)\theta*, c) \ge W^{\alpha}(\theta,c)\, W^{1-\alpha}(\theta*,c).$$

(ii) $W(\sum_1^m \alpha_i\, \theta^{(i)}, c) \ge W(\theta^{(1)}, c)$, if $\alpha_i \ge 0$, $\sum_1^m \alpha_i = 1$ and

$$W(\theta^{(1)},c) = W(\theta^{(2)},c) = \ldots = W(\theta^{(m)},c).$$

(iii) $W(\rho\theta,c) \ge W(\theta,c)$, for $|\rho| \le 1$.

Proof of Lemma 2.1

(i) From equation (2.9), the function $W(\theta,c)$ may also be written as

$$W(\theta,c) = \int_{\mathbf{R}^k} g^*(\theta,\mathrm{y})\, dy , \tag{2.15}$$

where the function $g^*(\theta,\mathrm{y})$ is log-concave in $(\theta,\mathrm{y})\in\mathbf{R}^{2k}$ (see, for example, Eaton (1987) p.79) and is given by

$$g^*(\theta,y) = I_\omega(y)\; g(y - \theta)$$

where $I_\omega(y)$ is the indicator function of the convex set ω. Then (i) follows from the equation (2.15) and Theorem A2 in the appendix.

(ii) By noting the condition that $W(\theta^{(1)},c) = W(\theta^{(2)},c) = \ldots = W(\theta^{(m)},c)$, (ii) follows from (i) by induction on m.

(iii) Notice that since $\frac{1+\rho}{2}\theta + \frac{1-\rho}{2}(-\theta) = \rho\theta$, property (ii) above and (2.7) imply (iii) #

Now we are in the position to give the main results. First, Theorem 2.2 shows that for the difference zone $\Theta^1_{b_0}$, the least favourable configuration of population means for the Studentised range test has exactly $k-1$ means equal, with the remaining mean chosen so that $b_1(\mu) = b_0\,\sigma$.

Theorem 2.2 Let $\mu_1^* = (0, \ldots, 0, kb_0\sigma/(k-1)) \in \mathbf{R}^k$. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu\in\Theta^1_{b_0}} \beta_S(\mu/\sigma, \nu,\alpha) = \beta_S(\mu_1^*/\sigma, \nu,\alpha).$$

Proof of Theorem 2.2

To prove the theorem, we only need to prove that for any real number $c \geq 0$ and $b \geq 0$

$$b_1(\theta) \equiv \max_{1\leq i\leq k} |\theta_i - \bar\theta| \geq b \quad \Rightarrow \quad W(\theta,c) \leq W(\theta_1^*,c),$$

where $\theta_1^* = (0, \ldots, 0, kb/(k-1))$. Without loss of generality, we suppose that $b_1(\theta) = \theta_k - \bar\theta = \tilde b \geq b$. Let $\theta^{(i)}$, $i = 1,\ldots,(k-1)!$ denote the vectors obtained from θ by permuting the first $k-1$ coordinates $\theta_1, \ldots, \theta_{k-1}$ and leaving the last coordinate θ_k in place. Let $\bar\theta_k = \frac{1}{k-1}\sum_{i=1}^{k-1}\theta_i$, so that

$$\theta_k - \bar\theta_k = \frac{k}{k-1}(\theta_k - \bar\theta) = \frac{k}{k-1}\tilde b.$$

Now by properties (2.7b) and Lemma 2.1, it follows that for any $c\in\mathbf{R}$

$$W(\theta,c) \leq W\Big(\frac{1}{(k-1)!}\sum_{i=1}^{(k-1)!}\theta^{(i)}, c\Big)$$

$$= W((\bar\theta_k, \ldots, \bar\theta_k, \theta_k), c)$$

$$= W(\ (0,\ \ldots,\ 0,\ \theta_k - \bar{\theta}_k),\ c\ )$$

$$= W(\ (0,\ldots,\ 0,\ \frac{k}{k-1}\tilde{b}),\ c\ )$$

$$\leq W(\ (0,\ldots,\ 0,\ \frac{k}{k-1}b),\ c\ )$$

$$= W(\ \theta_1^*,\ c\ )\ .$$

This is exactly what we want. So, the proof is completed #

The following theorem, Theorem 2.3, shows that the least favourable configurations of the F-test for the difference zone $\Theta_{b_0}^1$ are exactly as those of the Studentised range test.

Theorem 2.3 Let $\mu_1^* = (0,\ \ldots\ ,\ 0,\ kb_0\sigma/(k-1)) \in \mathbf{R}^k$ as in Theorem 2.2. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu \in \Theta_{b_0}^1} \beta_F(\mu/\sigma,\ \nu,\alpha) = \beta_F(\mu_1^*/\sigma,\ \nu,\alpha).$$

Proof of Theorem 2.3

It follows from equation (2.2) and the well known fact that $P\ (\ F_{k-1,\nu}\ (\delta) \geq F_{k-1,\nu}^{\alpha}\ )$ increases as δ increases, that in order to prove Theorem 2.3 it is necessary only to show that, under the condition $b_1(\mu) \geq b_0\sigma$,

$$\delta^2 = \frac{n}{\sigma^2} \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2$$

attains its minimum value at μ_1^*. It is easy to see this minimum will be attained at some μ which satisfies $b_1(\mu) = b_0\sigma$. So, the problem becomes to minimise $\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2$ under the condition $b_1(\mu) = b_0\sigma$. This is easy to establish by calculus methods, and we do not fill in the details here #

Now we turn our attention to the difference zone $\Theta_{b_0}^2$. Theorem 2.4 shows that for this difference zone, the least favourable configuration of population means for the Studentised range test has $k-2$ means chosen to be 0, with the other two means chosen to be $-b\sigma/2$ and $b\sigma/2$ respectively.

Theorem 2.4 Let $\mu_2^* = (-b_0\sigma/2, 0, \ldots, 0, b_0\sigma/2) \in \mathbf{R}^k$. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu \in \Theta_{b_0}^2} \beta_S(\mu/\sigma, \nu, \alpha) = \beta_S(\mu_2^*/\sigma, \nu, \alpha).$$

Proof of Theorem 2.4

If we can prove that for any real number $c \geq 0$ and $b \geq 0$

$$b_2(\theta) \equiv \max_{1 \leq i,j \leq k} |\theta_i - \theta_j| \geq b \quad \Rightarrow \quad W(\theta, c) \leq W(\theta_2^*, c),$$

where $\theta_2^* = (-b/2, 0, \ldots, 0, b/2)$, then the theorem will follows. Without loss of generality, we assume that $b_2(\theta) = \theta_k - \theta_1 = \tilde{b} \geq b$. Let $\theta^{(i)}$, $i = 1, \ldots, (k-2)!$ denote the vectors obtained from θ by permuting the middle $k-2$ coordinates $\theta_2, \ldots, \theta_{k-1}$ and leaving the first coordinate, θ_1, and the last coordinate, θ_k, in place. Let $\bar{\theta}_{1k} = \frac{1}{k-2}\sum_{i=2}^{k-1} \theta_i$. Then by properties (2.7b) and Lemma 2.1, it follows that for any $c \in \mathbf{R}$,

$$\begin{aligned}
W(\theta, c) &\leq W\Big(\frac{1}{(k-2)!}\sum_{1}^{(k-2)!} \theta^{(i)}, c\Big) \\
&= W((\theta_1, \bar{\theta}_{1k}, \ldots, \bar{\theta}_{1k}, \theta_k), c) \\
&\leq W\Big(\Big\{\frac{1}{2}(\theta_1, \bar{\theta}_{1k}, \ldots, \bar{\theta}_{1k}, \theta_k) + \\
&\qquad \frac{1}{2}(-\theta_k, -\bar{\theta}_{1k}, \ldots, -\bar{\theta}_{1k}, -\theta_1)\Big\}, c\Big) \\
&= W\Big(\Big(-\frac{1}{2}\tilde{b}, 0, \ldots, 0, \frac{1}{2}\tilde{b}\Big), c\Big) \\
&\leq W\Big(\Big(-\frac{1}{2}b, 0, \ldots, 0, \frac{1}{2}b\Big), c\Big) \\
&= W(\theta_2^*, c).
\end{aligned}$$

Thus, the proof is finished #

The following theorem shows that the F-test has exactly the same least favourable configurations as those of the Studentised range test for the difference zone $\Theta_{b_0}^2$.

Theorem 2.5 Let $\mu_2^* = (-b_0\sigma/2, \ldots, 0, b_0\sigma/2) \in \mathbf{R}^k$, as in Theorem 2.4. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu \in \Theta_{b_0}^2} \beta_F(\mu/\sigma, \nu, \alpha) = \beta_F(\mu_2^*/\sigma, \nu, \alpha).$$

Proof of Theorem 2.5

For the same reasons as in the proof of Theorem 2.3, in order to prove Theorem 2.5 it is necessary only to show that under the condition $b_2(\mu) \geq b_0\sigma$,

$$\delta^2 = \frac{n}{\sigma^2} \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2$$

attains its minimum value at μ_2^*. Again, this problem is obviously equivalent to minimising $\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2$ under the condition $b_2(\mu) = b_0\sigma$. This is easy to establish, and we omit the details here #

Theorem 2.2 to Theorem 2.5 have solved the problem of the least favourable configurations of the Studentised range test and F-test for the difference zones $\Theta_{b_0}^1$ and $\Theta_{b_0}^2$. For the difference zone

$$\Theta_{b_0}^3 \equiv \{ \mu : b_3(\mu) \equiv \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2 \geq b_0\sigma \} , \tag{2.16}$$

it is obvious that the least favourable configurations of the F-test are $\{ \mu : b_3(\mu) = b_0\sigma \}$, but for the Studentised range test, this problem is not as yet clear. Kunert (1988) proved that for $k=3$, the least favourable configurations of the Studentised range test for $\Theta_{b_0}^3$ satisfy two of the population means are equal, and the third mean is chosen to meet the condition $b_3(\mu) = b_0\sigma$. An interesting result of Wald (1942) shows that for all tests of the null hypothesis H_0 which have the same size, the F-test maximises the average power over "sphere" surfaces $\{ \mu : b_3(\mu) = b \}$, $b > 0$, on which $b_F(\mu)$ is constant. So, if we define Ψ to be a family of tests of size α with the power function of each test satisfying :

power function ($\lambda\mu$) $\geq$ power function (μ)

for $\lambda \geq 1$, then the F-test would be the maximin test in the test family Ψ for the difference zone $\Theta_{b_0}^3$.

In the following section we use the results of this section to show how to design an experiment to meet certain power requirements.

2.3 Design of experiments

In this section we discuss how to design an experiment which guarantees a certain level of power against a specified set of "alternative" population means $\mu = (\mu_1, \cdots, \mu_k)$. One way to do this is to specify a function $b(\mu) \geq 0$ which measures the "variability" of the population means μ_i, $1 \leq i \leq k$. Examples of such functions are the $b_1(\mu)$, $b_2(\mu)$ and $b_3(\mu)$ defined in equations (2.13), (2.14) and (2.16). The experimenter then specifies a positive constant $b_0 > 0$ and a power level β, $\alpha \leq \beta \leq 1$, and requires that a test of size α of the null hypothesis H_0 (that the k population means are all equal), have power no less than β whenever $b(\mu) \geq \sigma b_0$, i.e.

$$b(\mu) \geq \sigma b_0 \qquad \Rightarrow \qquad \beta(\mu,\sigma,\nu,\alpha) \geq \beta, \tag{2.17}$$

where $\beta(\mu,\sigma,\nu,\alpha)$ is the power function of the test procedure being used. The problem of experimental design is then to find the smallest sample size n for which (2.17) holds. This is guaranteeing the power value of the test to be no less than β for the population means in the difference zone $\Theta_{b_0} \equiv \{ \mu : b(\mu) \geq b_0\sigma \}$; this can be established by finding the smallest sample size n such that $\beta(\mu,\sigma,\nu,\alpha) \geq \beta$ at the least favourable configuration of population means for the difference zone Θ_{b_0}.

For the F-test it is natural to use the measure function $b(\mu) = b_3(\mu)$, given in equation (2.16), since the power function $\beta_F(\mu/\sigma, \nu,\alpha)$ depends on the population means μ only through the quantity $b_3(\mu)$. Notice that the F-test is the maximin test in the test family Ψ (defined at the end of last section) for the difference zone $\Theta^3_{b_0}$, so we should use the F-test preferentially if we want to guarantee the power level for the difference zone $\Theta^3_{b_0}$. The least favourable configuration in this case is the set of all vectors of population means μ such that $b_F(\mu) = b_0\sigma$; tables of the power function have been calculated which allow the necessary sample size n to be determined in order to guarantee the probability requirements.

However, it seems more intuitive to use $b_1(\mu)$ and $b_2(\mu)$ as the measurements of the variability of the population means μ; since neither the F-test nor the Studentised range test has been proved to be the maximin test for the difference zone $\Theta_{b_0}^1$ or $\Theta_{b_0}^2$ it is of great interest to compare the least favourable power of the F-test and the Studentised range test for both the difference zones in order to assess the merits of these two tests in these cases. In the following section, we begin by comparing the least favourable power of the F-test and the Studentised range test for the difference zone $\Theta_{b_0}^1$.

From Theorem 2.2 and Theorem 2.3, the F-test and the Studentised range test attain their least favourable power in the difference zone $\Theta_{b_0}^1$ at the same configuration of the population means, i.e. the least favourable configuration $\mu_1^* = (0, \ldots, 0, kb_0\sigma/(k-1))$ where $b_0 \geq 0$. So, the least favourable power of the Studentised range test for the difference zone $\Theta_{b_0}^1$, $\beta_S(\mu_1^*/\sigma, \nu, \alpha)$ can be calculated from equation (2.4) with

$$W(\sqrt{n}\mu_1^*/\sigma, \sqrt{s}q_{k,\nu}^{\alpha}) = P\left[|Y_i - Y_j| \leq \sqrt{s}q_{k,\nu}^{\alpha}\ ,\ 1 \leq i,j \leq k\right]$$

$$= P\left[Y_k \leq Y_1, \ldots, Y_{k-1} \leq Y_k + \sqrt{s}q_{k,\nu}^{\alpha}\right]$$
$$+ (k-1)\, P\left[Y_1 \leq Y_2, \ldots, Y_k \leq Y_1 + \sqrt{s}q_{k,\nu}^{\alpha}\right]$$

$$= \int_{-\infty}^{\infty} \varphi(x - b_0 k\sqrt{n}/(k-1))\ [\Phi(x+\sqrt{s}q_{k,\nu}^{\alpha}) - \Phi(x)]^{k-1}\, dx$$
$$+ (k-1)\int_{-\infty}^{\infty} \varphi(x)[\Phi(x+\sqrt{s}q_{k,\nu}^{\alpha}) - \Phi(x)]^{k-2} \times$$
$$[\Phi(x+\sqrt{s}q_{k,\nu}^{\alpha} - b_0 k\sqrt{n}/(k-1)) - \Phi(x - b_0 k\sqrt{n}/(k-1))]\, dx\ , \qquad (2.18)$$

where the Y_i are independent random variables with the following distributions

$$Y_1, \ldots, Y_{k-1} \sim N(0,1), \quad Y_k \sim N(bk\sqrt{n}/(k-1), 1),$$

and $\varphi(\cdot)$ and $\Phi(\cdot)$ are respectively the p.d.f. and c.d.f. of the standard normal distribution. Equation (2.18) with $\nu = k(n-1)$ allows us to calculate the least favourable power of the Studentised range test for the difference zone $\Theta_{b_0}^1$

for specified b_0, α and n. In fact, the calculation involves only a two dimensional numerical integral by using the NAG routine for $\Phi(\cdot)$.

Table 2.1 and Table 2.2 present some examples of the least favourable power level for $b_0=0.5$ and $b_0=1.0$ respectively, for $\alpha=0.05$, $k=3, \ldots, 8$ and a selection of sample sizes n between 5 and 50. For example, if a power level of at least 0.90 is required for an experiment with $k=3$ populations, then over 35 observations are required from each of the three populations with $b_0=0.5$, but 10 observations from each of the three populations will suffice with $b_0=1.0$. Unfortunately, many of the critical points $q^{\alpha}_{k,\nu}$ of the Studentised range are not tabulated, and so in the preparation of Table 2.1 and Table 2.2 it was first necessary to calculate the critical points required. A computer program has been developed which enables these power calculations to be made for general b, α, k, and n.

The calculation of the least favourable power of the F-test for the difference zone $\Theta^1_{b_0}$ can also be done by a two dimensional numerical integral. There are, in fact, many calculation methods available in the literature, see for example Tiku (1967). It is interesting to compare the power levels of the Studentised range test and the F-test under their common least favourable configuration of population means for the difference zone $\Theta^1_{b_0}$. Such a comparison is presented in Table 2.3 for $\alpha=0.05$ and certain values of k, b_0, and n; the values of the power of the F-test being taken from Tiku (1967). Clearly there is very little difference between the two power functions, the only differences of any magnitude being for "intermediate" values of the parameter b_0 since when $b_0=0$ both power functions are equal to α, and for large b_0 both power functions approach 1. The indications are that for $k=3$ and 4, the F-test is very slightly more powerful, but for larger numbers of populations k, the Studentised range test is more powerful. In fact, for $\alpha=0.05$, $k=10$, and $n=7$, the power of the Studentised range test may exceed the power of the F-test by as much as 0.07.

We now turn our attention to the difference zone $\Theta^2_{b_0}$. Pearson & Hartley (1951) considered the upper and lower bounds of the range $\max_{1\le i\le} |\mu_i-\mu_j|$ when the non-central parameter of the F-test, $\delta^2 = \frac{n}{\sigma^2}\sum_{i=1}^{k}(\mu_i-\bar{\mu})^2$ was fixed, and then used these bounds as indications of the differences among the treatment means for that specific δ^2 value. The upper bound for different δ^2 has been tabulated in

Kastenbaum, Hole & Bowman (1970), together with the respective power value of the F-test. A computational investigation of the power function of the Studentised range test at what is in fact the least favourable configuration for $\Theta^2_{b_0}$ has been performed before by David, Lachenbruch & Brandis (1972), but a proof of the least favourable configuration has not been given. Our Theorem 2.4 and Theorem 2.5 show that for the difference zone $\Theta^2_{b_0}$, the least favourable configuration of the Studentised range test and the F-test have the same form, $\mu_2^* = (-b_0\sigma/2, 0, \ldots, 0, b_0\sigma/2)$. So, the least favourable power of the Studentised range test $\beta_S(\mu_2^*/\sigma, \nu, \alpha)$ may be calculated from equation (2.4) with

$$W(\sqrt{n}\mu_2^*/\sigma, \sqrt{s}q^{\alpha}_{k,\nu}) = P\left[|Y_i - Y_j| \le \sqrt{s}q^{\alpha}_{k,\nu} \ , \ 1 \le i,j \le k \right]$$

$$= P\left[Y_k \le Y_1, \ldots, Y_{k-1} \le Y_k + \sqrt{s}q^{\alpha}_{k,\nu} \right]$$
$$+ (k-2)\, P\left[Y_2 \le Y_1, Y_3, \ldots, Y_k \le Y_2 + \sqrt{s}q^{\alpha}_{k,\nu} \right]$$
$$+ P\left[Y_1 \le Y_2, \ldots, Y_k \le Y_1 + \sqrt{s}q^{\alpha}_{k,\nu} \right]$$

$$= \int_{-\infty}^{\infty} \varphi(x - b_0\sqrt{n}/2)\ [\Phi(x + \sqrt{s}q^{\alpha}_{k,\nu}) - \Phi(x)]^{k-2} \times$$
$$[\Phi(x + \sqrt{s}q^{\alpha}_{k,\nu} + b_0\sqrt{n}/2) - \Phi(x + b_0\sqrt{n}/2)]\ dx$$

$$+ (k-2) \int_{-\infty}^{\infty} \varphi(x)\, [\Phi(x + \sqrt{s}q^{\alpha}_{k,\nu}) - \Phi(x)]^{k-3} \times$$
$$[\Phi(x + \sqrt{s}q^{\alpha}_{k,\nu} - b_0\sqrt{n}/2) - \Phi(x - b_0\sqrt{n}/2)] \times$$
$$[\Phi(x + \sqrt{s}q^{\alpha}_{k,\nu} + b_0\sqrt{n}/2) - \Phi(x + b_0\sqrt{n}/2)]\ dx$$

$$+ \int_{-\infty}^{\infty} \varphi(x + b_0\sqrt{n}/2)\ [\Phi(x + \sqrt{s}q^{\alpha}_{k,\nu}) - \Phi(x)]^{k-2} \times$$

$$[\Phi(x+\sqrt{s}\,q^{\alpha}_{k,\nu}-b_0\sqrt{n}/2)-\Phi(x-b_0\sqrt{n}/2)]\ dx\ , \tag{2.19}$$

where the Y_i are independent random variables with the following distributions

$$Y_1 \sim N(-b_0\sqrt{n}/2,1), \quad Y_2, \ldots, Y_{k-1} \sim N(0,1), \quad Y_k \sim N(b_0\sqrt{n}/2,1).$$

Equation (2.19) with $\nu=k(n-1)$ allows us to calculate the power of the Studentised range test at the least favourable configuration μ_2^* for specified b_0, α and n. Again this calculation involves only a two dimensional numerical integral by using the NAG routine for $\Phi(\cdot)$.

Table 2.4 and Table 2.5 present some examples of this power level for $b_0=0.5$ and $b_0=1.0$ respectively, for $\alpha=0.05$, $k=3, \ldots, 8$ and a selection of sample sizes n between 5 and 50. Again, a computer program has been developed which enables these power calculations to be made for general b_0, α, k, and n.

Again we can compare the power levels of the Studentised range test and the F-test under their common least favourable configuration for the difference zone $\Theta^2_{b_0}$. This comparison is presented in Table 2.6 for $\alpha=0.05$ and certain values of k, b_0, and n, the values of the power of the F-test being taken from Tiku (1967). The indications are similar to those of Table 3 in that for $k=3$ and 4, the F-test is very slightly more powerful, but for larger numbers of populations k, the Studentised range test is more powerful.

In this section, since the experiment is designed as a one sample procedure, it is necessary to state the difference zones $\Theta^1_{b_0}$ and $\Theta^2_{b_0}$ in terms of the unknown variance σ. This may be avoided only by using a two-stage sample procedure in which examination of the first sample indicates what sample sizes are required in the second stage to guarantee the probability requirements. The next section is devoted to the study of some two-stage sample procedures.

2.4 Two-stage sample test procedures

In this section we consider the problem of guaranteeing the power for the difference zones

$$\tilde{\Theta}_{b_0}^1 \equiv \{ \mu : b_1(\mu)=\max_{1\le i\le k}|\mu_i-\bar{\mu}|\ge b_0 \}$$

and

$$\tilde{\Theta}_{b_0}^2 \equiv \{ \mu : b_2(\mu)=\max_{1\le i,j\le k}|\mu_i-\mu_j|\ge b_0 \}.$$

In many practical situations, an experimenter may want to guarantee the power for the difference zones $\tilde{\Theta}_{b_0}^1$ or $\tilde{\Theta}_{b_0}^2$ rather than $\Theta_{b_0}^1$ or $\Theta_{b_0}^2$, since in the latter two the nuisance parameter σ is involved. For this purpose, a fixed sample size test procedure will not work, because for a sample of fixed size there does not exist a nontrivial test procedure for the null hypothesis $H_0 : \mu_1=\ldots=\mu_k$ whose power is independent of the nuisance parameter σ (see Stein (1945)). To cope with this difficulty, two-stage sample simultaneous test procedures are needed. Healy (1956) studied some two-stage sample procedures mainly for the purpose of obtaining simultaneous confidence intervals by developing the ideas of Stein (1945). In the following, we first derive two two-stage sample test procedures whose power functions depend only on $\mu = (\mu_1, \ldots, \mu_k)$, find out their least favourable configurations for the difference zone $\tilde{\Theta}_{b_0}^1$ and $\tilde{\Theta}_{b_0}^2$, and the method of calculating their respective least favourable powers. Then we introduce two other two-stage sample test procedures whose power functions, although dependent on the nuisance parameter σ, dominate the power functions of the previous two test procedures. As we will see, the least favourable configurations of the latter two test procedures for the difference zones $\tilde{\Theta}_{b_0}^1$ and $\tilde{\Theta}_{b_0}^2$ are the same as those of the previous two test procedures, thus a lower bound on the least favourable power of the latter two test procedures can be obtained as the least favourable power of the previous two.

We still consider the one-way fixed effects analysis of variance model (2.1)

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad 1\le i\le k, \quad j=1, 2, \ldots..$$

Notice that instead of taking one sample of fixed size n from each of the k populations, this time the total sample size we take from each population is a random variable. At the first stage, we take n_0 (non-random) observations from each population, and obtain an estimate, S^2, of the variance σ^2 by

$$S^2 = \frac{1}{\nu}\sum_{i=1}^{k}\sum_{j=1}^{n_0}(X_{ij}-\bar{X}_i)^2, \quad \bar{X}_i = \frac{1}{n_0}\sum_{j=1}^{n_0}X_{ij}, \quad 1\le i\le k \tag{2.20}$$

where $\nu=k(n_0-1)$. Define N by

$$N = \max\{\ [\frac{S^2}{z}]+1, n_0+1\ \}\ , \tag{2.21}$$

where z is some previously specified positive constant and $[a]$ denotes the largest integer less than or equal to a. Then at the second stage we take a sample of size $N-n_0$ from each population

$$X_{i\,n_0+1}\ ,\ X_{i\,n_0+2}\ ,\ \dots\ ,\ X_{i\,N}\ , \qquad 1\le i\le k,$$

and choose real numbers a_i, $i=1, \dots, N$ in such a way that

$$a_1 = a_2 = \dots = a_{n_0}\ ,$$

$$\sum_{i=1}^{N} a_i = 1\ ,$$

$$S^2 \sum_{i=1}^{N} a_i^2 = z\ .$$

This is clearly possible since $N \ge n_0 + 1$ and

$$\min \sum_{i=1}^{N} a_i^2 = \frac{1}{N} \le \frac{z}{S^2} \quad \text{by (2.21)}\ ,$$

the minimum being taken subject to the conditions

$$\sum_{i=1}^{N} a_i = 1, \quad a_1 = a_2 = \dots = a_{n_0}.$$

Denote

$$\eta_i = \frac{\sum_{j=1}^{N} a_j X_{ij} - \mu_i}{\sqrt{z}} = \frac{\sum_{j=1}^{N} a_j (X_{ij} - \mu_i)}{\sqrt{z}}, \quad 1\le i\le k\ . \tag{2.22}$$

The basis of this sampling procedure is the following lemma, which gives the joint distribution of $\eta = (\eta_1, \eta_2, \dots, \eta_k)$.

Lemma 2.2 For the above sampling procedure, η has the same joint distribution as

$$\left(\frac{\xi_1}{\sqrt{\frac{S^2}{\sigma^2}}}, \frac{\xi_2}{\sqrt{\frac{S^2}{\sigma^2}}}, \dots , \frac{\xi_k}{\sqrt{\frac{S^2}{\sigma^2}}} \right)$$

where $\xi_1, \dots, \xi_k$ are i.i.d. N(0, 1) random variables independent of $\frac{S^2}{\sigma^2}$ as defined in (2.20) which has a χ^2_ν/ν distribution.

Proof of Lemma 2.2

Under the condition that S^2 is given, we have

$$\frac{\xi_i}{\sqrt{S^2/\sigma^2}}, \quad 1 \le i \le k \quad \text{are} \quad i.i.d. \ N(0, \frac{1}{S^2/\sigma^2}).$$

In order to prove the lemma, we only need to show that the same statement is also true for $\eta = (\eta_1, \ldots, \eta_k)$, i.e. under the condition that S^2 is given :

$$\eta_i \ (1 \le i \le k) \quad \text{are} \quad i.i.d. \ N(0, \frac{1}{S^2/\sigma^2}).$$

Notice that under the condition that S^2 is given, N and a_i $(1 \le i \le N)$ are then also fixed. So, it is easy to see from (2.22) that the η_i $(1 \le i \le k)$ are independent and normally distributed with

$$E\,\eta_i = 0, \quad Var\,\eta_i = \frac{1}{z}\sum_{j=1}^{N} a_j^2\sigma^2 = \sigma^2/S^2.$$

The proof is thus finished #

Now we investigate two test procedures of the null hypothesis H_0, which are based on the above sampling procedure. First, the range type test procedure operates as follows. Define

$$Q = \frac{\max\limits_{1 \le i,j \le k} \left| \sum\limits_{r=1}^{N} a_r X_{ir} - \sum\limits_{s=1}^{N} a_s X_{js} \right|}{\sqrt{z}}.$$

A size α test is then obtained by rejecting the null hypothesis H_0 iff

$$Q \ge q_{k,\nu}^{\alpha}, \tag{2.23}$$

where $q_{k,\nu}^{\alpha}$ is the upper α-point of the Studentised range distribution with parameters k and ν. Since under the null hypothesis H_0 :

$$Q = \frac{\max\limits_{1 \le i,j \le k} \left| \sum\limits_{r=1}^{N} a_r X_{ir} - \sum\limits_{s=1}^{N} a_s X_{js} \right|}{\sqrt{z}} = \max_{1 \le i,j \le k} | \eta_i - \eta_j |,$$

this, as follows from Lemma 2.2, has a Studentised range distribution with parameters k and ν. So, the test defined by (2.23) has size exactly equal to α. The power of this test procedure is

$$\beta_Q(\mu, n_0, z) = P\{\ Q \ge q_{k,\nu}^{\alpha}\ \}$$

$$= P\{ \frac{\max\limits_{1\le i,j\le k} | \sum\limits_{r=1}^{N} a_r X_{ir} - \sum\limits_{s=1}^{N} a_s X_{js} |}{\sqrt{z}} \ge q_{k,\nu}^{\alpha} \}$$

$$= P\{ \max_{1\le i,j\le k} | \frac{(\sum\limits_{r=1}^{N} a_r X_{ir} - \mu_i) - (\sum\limits_{s=1}^{N} a_s X_{js} - \mu_j)}{\sqrt{z}} + \frac{\mu_i - \mu_j}{\sqrt{z}} | \ge q_{k,\nu}^{\alpha} \}$$

$$= P\{ \max_{1\le i,j\le k} | \eta_i - \eta_j + \frac{\mu_i - \mu_j}{\sqrt{z}} | \ge q_{k,\nu}^{\alpha} \}$$

$$= P\{ \max_{1\le i,j\le k} | \frac{\xi_i - \xi_j}{\sqrt{S^2/\sigma^2}} + \frac{\mu_i - \mu_j}{\sqrt{z}} | \ge q_{k,\nu}^{\alpha} \}$$

$$= 1 - P\{ \max_{1\le i,j\le k} | \frac{\xi_i - \xi_j}{\sqrt{S^2/\sigma^2}} + \frac{\mu_i - \mu_j}{\sqrt{z}} | \le q_{k,\nu}^{\alpha} \}$$

$$= 1 - \int_{s=0}^{\infty} P\{ \max_{1\le i,j\le k} | \frac{\xi_i - \xi_j}{\sqrt{s}} + \frac{\mu_i - \mu_j}{\sqrt{z}} | \le q_{k,\nu}^{\alpha} \} f_{\chi_\nu^2/\nu}(s)\, ds \qquad (2.24)$$

where the fifth equation above follows from Lemma 2.2, and $f_{\chi_\nu^2/\nu}(s)$ is the probability density function of a χ_ν^2/ν random variable. So, the power function $\beta_Q(\mu, n_0, z)$ is clearly independent of the nuisance parameter σ^2. Furthermore, notice that

$$P\{ \max_{1\le i,j\le k} | \frac{\xi_i - \xi_j}{\sqrt{s}} + \frac{\mu_i - \mu_j}{\sqrt{z}} | \le q_{k,\nu}^{\alpha} \}$$

$$= P\{ \max_{1\le i,j\le k} | \xi_i - \xi_j + \frac{\sqrt{s}}{\sqrt{z}}(\mu_i - \mu_j) | \le \sqrt{s}\, q_{k,\nu}^{\alpha} \}$$

$$= W(\theta, c) \ , \qquad (2.25)$$

where $W(\theta, c)$ is defined in (2.5) and (2.6) with

$$\theta_i = \sqrt{\frac{s}{z}}\, \mu_i \quad 1\le i\le k, \quad c = \sqrt{s}\, q_{k,\nu}^{\alpha} \ . \qquad (2.26)$$

We have the following theorem about the least favourable configuration of this test procedure for the difference zones $\tilde{\Theta}_{b_0}^1$ and $\tilde{\Theta}_{b_0}^2$.

Theorem 2.6

(i) Let $\tilde{\mu}_1 = (0,\ldots,0,kb_0/(k-1)) \in \mathbf{R}^k$. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu \in \tilde{\Theta}^1_{b_0}} \beta_Q(\mu,n_0,z) = \beta_Q(\tilde{\mu}_1,n_0,z).$$

(ii) Let $\tilde{\mu}_2 = (-b_0/2,0,\ldots,0,b_0/2) \in \mathbf{R}^k$. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu \in \tilde{\Theta}^2_{b_0}} \beta_Q(\mu,n_0,z) = \beta_Q(\tilde{\mu}_2,n_0,z).$$

Proof of Theorem 2.6

The proof follows immediately from equations (2.24), (2.25), (2.26) and the proofs of Theorem 2.2 and Theorem 2.4. The details are omitted here #

The calculation of the least favourable power of the test procedure for the difference zones $\tilde{\Theta}^1_{b_0}$ and $\tilde{\Theta}^2_{b_0}$, $\beta_Q(\tilde{\mu}_1,n_0,z)$ and $\beta_Q(\tilde{\mu}_2,n_0,z)$ can be carried out following equations (2.24) and (2.25) with $W(\theta,c)$ being expressed in the same way as in equation (2.18) or (2.19) respectively. So, again a two dimensional numerical integral is involved.

As with the one sample case, an alternative test procedure, a two-stage sample F type test, can be derived and which operates as follows. Define

$$L = \frac{1}{(k-1)} \sum_{i=1}^{k} \left(\frac{1}{\sqrt{z}} \sum_{j=1}^{N} a_j X_{ij} - \frac{1}{k\sqrt{z}} \sum_{l=1}^{k} \sum_{j=1}^{N} a_j X_{lj}\right)^2 .$$

Then a size α test based on L is to reject the null hypothesis H_0 iff

$$L \geq F^{\alpha}_{k-1,\nu} ,$$

where $F^{\alpha}_{k-1,\nu}$ is the upper α-point of the F distribution with parameters $k-1$ and $\nu = k(n_0-1)$. This test has size exactly equal to α since under the null hypothesis H_0

$$L = \frac{1}{(k-1)} \sum_{i=1}^{k} (\eta_i - \bar{\eta})^2 ,$$

which, as follows from Lemma 2.2, has a F distribution with parameters k and ν. The power of this test procedure, $\beta_L(\mu,n_0,z)$ satisfies

$$1 - \beta_L(\mu,n_0,z) = P\{ L \leq F^{\alpha}_{k-1,\nu} \}$$

$$= P\{ \frac{1}{(k-1)} \sum_{i=1}^{k} \left(\frac{1}{\sqrt{z}} \sum_{j=1}^{N} a_j X_{ij} - \frac{1}{k\sqrt{z}} \sum_{l=1}^{k} \sum_{j=1}^{N} a_j X_{lj}\right)^2 \leq F^{\alpha}_{k-1,\nu} \}$$

$$= P\{ \frac{1}{(k-1)} \sum_{i=1}^{k} \left[\frac{1}{\sqrt{z}} \left(\sum_{j=1}^{N} a_j X_{ij} - \mu_i + \mu_i\right) - \right.$$

$$\frac{1}{k\sqrt{z}}(\sum_{l=1}^{k}\sum_{j=1}^{N}a_jX_{lj}-\mu_l+\mu_l)]^2 \le F^{\alpha}_{k-1,\nu} \}$$

$$= P\{ \frac{1}{(k-1)} \sum_{i=1}^{k} [(\eta_i+\frac{\mu_i}{\sqrt{z}}) - \frac{1}{k}\sum_{l=1}^{k}(\eta_l+\frac{\mu_l}{\sqrt{z}})]^2 \le F^{\alpha}_{k-1,\nu} \}$$

$$= P\{ \frac{1}{(k-1)} \sum_{i=1}^{k} [(\frac{\xi_i}{\sqrt{S^2/\sigma^2}}+\frac{\mu_i}{\sqrt{z}}) - \frac{1}{k}\sum_{l=1}^{k}(\frac{\xi_l}{\sqrt{S^2/\sigma^2}}+\frac{\mu_l}{\sqrt{z}})]^2 \le F^{\alpha}_{k-1,\nu} \}$$

$$= \int_{s=0}^{\infty} P\{ \frac{1}{(k-1)} \sum_{i=1}^{k} [(\frac{\xi_i}{\sqrt{s}}+\frac{\mu_i}{\sqrt{z}}) - \frac{1}{k}\sum_{l=1}^{k}(\frac{\xi_l}{\sqrt{s}}+\frac{\mu_l}{\sqrt{z}})]^2 \le F^{\alpha}_{k-1,\nu} \}$$

$$\times f_{\chi^2_\nu/\nu}(s)\, ds \,, \tag{2.27}$$

where $f_{\chi^2_\nu/\nu}(s)$ is the probability density function of a χ^2_ν/ν random variable, and $\xi_1,\xi_2,\ldots,\xi_k$ are as defined in Lemma 2.2. So, the power function $\beta_L(\mu,n_0,z)$ is independent of the nuisance parameter σ. Furthermore, notice that

$$P\{ \frac{1}{(k-1)} \sum_{i=1}^{k} [(\frac{\xi_i}{\sqrt{s}}+\frac{\mu_i}{\sqrt{z}}) - \frac{1}{k}\sum_{l=1}^{k}(\frac{\xi_l}{\sqrt{s}}+\frac{\mu_l}{\sqrt{z}})]^2 \le F^{\alpha}_{k-1,\nu} \}$$

$$= P\{ \frac{1}{(k-1)} \sum_{i=1}^{k} [(\xi_i+\sqrt{\frac{s}{z}}\,\mu_i) - \frac{1}{k}\sum_{l=1}^{k}(\xi_l+\sqrt{\frac{s}{z}}\,\mu_l)]^2 \le s\, F^{\alpha}_{k-1,\nu} \}$$

$$= P\{ \frac{1}{k-1}\, \chi^2_{k-1}(\delta) \le s\, F^{\alpha}_{k-1,\nu} \} \,, \tag{2.28}$$

where $\chi^2_{k-1}(\delta)$ denotes a non-central χ^2 random variable with non-centrality parameter δ and degree of freedom $k-1$, and

$$\delta^2 = \frac{s}{z} \sum_{i=1}^{k} (\mu_i-\bar{\mu})^2 \quad \text{with} \quad \bar{\mu} = \frac{1}{k}\sum_{i=1}^{k}\mu_i \,. \tag{2.29}$$

We have the following results about the least favourable configurations of this test procedure for the difference zones $\tilde{\Theta}^1_{b_0}$ and $\tilde{\Theta}^2_{b_0}$.

Theorem 2.7

(i) Let $\tilde{\mu}_1 = (0,\ldots,0,kb_0/(k-1))\in\mathbf{R}^k$. Then for any real number $b_0\ge 0$,

$$\inf_{\mu\in\tilde{\Theta}^1_{b_0}} \beta_L(\mu,n_0,z) = \beta_L(\tilde{\mu}_1,n_0,z).$$

(ii) Let $\tilde{\mu}_2 = (-b_0/2,0,\ldots,0,b_0/2)\in\mathbf{R}^k$. Then for any real number $b_0\ge 0$,

$$\inf_{\mu \in \tilde{\Theta}^2_{b_0}} \beta_L(\mu, n_0, z) = \beta_L(\tilde{\mu}_2, n_0, z).$$

Proof of Theorem 2.7

Notice the fact that the right tail probability of a non-central χ^2 random variable increases when the non-centrality parameter increases. Theorem 2.7 follows immediately from equations (2.27), (2.28) and (2.29) #

The calculation of the least favourable power of the above test for the difference zones $\tilde{\Theta}^1_{b_0}$ and $\tilde{\Theta}^2_{b_0}$ can be done by using equations (2.27), (2.28) and (2.29), again involving only a two dimensional numerical integral. In fact

$$P\{ \frac{1}{k-1} \chi^2_{k-1}(\delta) \le s F^{\alpha}_{k-1,\nu} \}$$

$$= P\{ \chi^2_{k-2} + [N(\delta,1)]^2 \le (k-1)\, s F^{\alpha}_{k-1,\nu} \}$$

$$= \int_{x=0}^{\infty} P\{ (k-2)x + [N(\delta,1)]^2 \le (k-1)\, s F^{\alpha}_{k-1,\nu} \} f_{\chi^2_{k-2}/(k-2)}(x)\, dx$$

$$= \int_{x=0}^{A(s)} P\{ [N(\delta,1)]^2 \le B(x,s) \} f_{\chi^2_{k-2}/(k-2)}(x)\, dx$$

$$= \int_{x=0}^{A(s)} \{ \Phi(\delta + \sqrt{B(x,s)}) - \Phi(\delta - \sqrt{B(x,s)}) \} f_{\chi^2_{k-2}/(k-2)}(x)\, dx ,$$

where χ^2_{k-2} denotes a central χ^2 random variable with $k-2$ degrees of freedom, and

$$A(s) = \frac{(k-1)}{(k-2)}\, s F^{\alpha}_{k-1,\nu}, \quad B(x,s) = (k-1) s F^{\alpha}_{k-1,\nu} - (k-2)x.$$

So, by using the NAG routine for $\Phi(\,)$, the power calculation is only a two dimensional numerical integral.

In the above two test procedures, the first stage sample size, n_0, and, the constant z which controls the sample size of the second stage, must be chosen. Some specific guide lines for choosing n_0 are discussed in Moshman(1958) and Seelbinder(1953). For the chosen value of n_0, z is then chosen to guarantee the required power level for the difference zone $\tilde{\Theta}^1_{b_0}$ or $\tilde{\Theta}^2_{b_0}$.

In preceding section we have derived two two-stage sample simultaneous test procedures whose power functions are independent of the nuisance parameter σ, and solved the problem of guaranteeing the power value for these two test procedures for the difference zones $\tilde{\Theta}^1_{b_0}$ and $\tilde{\Theta}^2_{b_0}$. But the definitions of these two test procedures involve the real (random) numbers $a_1, a_2, \ldots, a_N$, whose choice is somewhat subjective, and make these tests not intuitively acceptable to an experimenter. Furthermore, these two tests waste information in order to make the power function strictly independent of σ. So in the following, we introduce two other test procedures, calling them the two-stage sample Studentised range test and the two-stage sample F-test both of which are more intuitive than the previous two tests. Although the powers of the two-stage sample Studentised range test and F-test both depend upon the nuisance parameter σ, their values are in fact no less than those of the previous two tests respectively. Healy (1956) used these two test statistics for obtaining simultaneous confidence intervals.

After drawing the first sample of size n_0 from each population and getting the estimate S^2 of the variance σ^2 as before, we define

$$\tilde{N} = \max\{ [\frac{S^2}{z}]+1, n_0 \} \tag{2.30}$$

instead of the N defined in (2.21). Then we draw a second sample of size $\tilde{N}-n_0$ from each population

$$X_{i\,n_0+1},\ X_{i\,n_0+2},\ \ldots\ ,\ X_{i\tilde{N}}, \qquad 1 \le i \le k,$$

and denote

$$\zeta_i = \frac{\sum_{j=1}^{\tilde{N}} (X_{ij}-\mu_i)}{\sqrt{\tilde{N}\, S^2}} .$$

We now have the following lemma analogous to Lemma 2.2, which gives the joint distribution of $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_k)$.

Lemma 2.3 For the above sampling procedure, ζ has the same joint distribution as

$$\left(\frac{\xi_1}{\sqrt{\frac{S^2}{\sigma^2}}}, \frac{\xi_2}{\sqrt{\frac{S^2}{\sigma^2}}}, \ldots , \frac{\xi_k}{\sqrt{\frac{S^2}{\sigma^2}}} \right)$$

where $\xi_1, \ldots, \xi_k$ are i.i.d. N(0, 1) random variables independent of $\frac{S^2}{\sigma^2}$ as defined in (2.20) which has a χ^2/ν distribution as before.

Proof of Lemma 2.3

As the proof is similar to that of Lemma 2.2, it is omitted #

Now we investigate the two-stage sample Studentised range test of the null hypothesis H_0, which operates as follows. Define

$$\tilde{Q} = \frac{\max_{1 \le i,j \le k} \left| \sum_{r=1}^{\tilde{N}} X_{ir} - \sum_{s=1}^{\tilde{N}} X_{js} \right|}{\sqrt{\tilde{N}\, S^2}} .$$

A size α test is obtained by rejecting the null hypothesis H_0 iff

$$\tilde{Q} \ge q^{\alpha}_{k,\nu} ,$$

where $q^{\alpha}_{k,\nu}$ has the same meaning as before. Since under the null hypothesis H_0

$$\tilde{Q} = \frac{\max_{1 \le i,j \le k} \left| \sum_{r=1}^{\tilde{N}} X_{ir} - \sum_{s=1}^{\tilde{N}} X_{js} \right|}{\sqrt{\tilde{N}\, S^2}} = \max_{1 \le i,j \le k} | \zeta_i - \zeta_j | ,$$

this, as follows from Lemma 2.3, has a Studentised range distribution with parameters k and ν. So, this test has size exactly equal to α. The power of this test procedure is

$$\beta_{\tilde{Q}}(\mu,\sigma,n_0,z) = P\{ \tilde{Q} \ge q^{\alpha}_{k,\nu} \}$$

$$= P\{ \frac{\max_{1 \le i,j \le k} \left| \sum_{r=1}^{\tilde{N}} X_{ir} - \sum_{s=1}^{\tilde{N}} X_{js} \right|}{\sqrt{\tilde{N}\, S^2}} \ge q^{\alpha}_{k,\nu} \}$$

$$= P\{ \frac{\max_{1 \le i,j \le k} \left| \sum_{r=1}^{\tilde{N}} (X_{ir} - \mu_i) - \sum_{s=1}^{\tilde{N}} (X_{js} - \mu_j) + \tilde{N}(\mu_i - \mu_j) \right|}{\sqrt{\tilde{N}\, S^2}} \ge q^{\alpha}_{k,\nu} \}$$

$$= P\{ \max_{1 \le i,j \le k} | \zeta_i - \zeta_j + \sqrt{\tilde{N}/S^2}(\mu_i - \mu_j) | \ge q^{\alpha}_{k,\nu} \}$$

$$= P\{ \max_{1 \le i,j \le k} \left| \frac{\xi_i - \xi_j}{\sqrt{S^2/\sigma^2}} + \sqrt{\tilde{N}/S^2}(\mu_i - \mu_j) \right| \ge q^{\alpha}_{k,\nu} \}$$

$$= 1 - P\{ \max_{1 \le i,j \le k} \left| \frac{\xi_i - \xi_j}{\sqrt{S^2/\sigma^2}} + \sqrt{\tilde{N}/S^2}(\mu_i - \mu_j) \right| \le q^{\alpha}_{k,\nu} \}$$

$$= 1 - \int_{s=0}^{\infty} P\{ \max_{1 \le i,j \le k} | \frac{\xi_i - \xi_j}{\sqrt{s}} + \sqrt{\frac{\tilde{N}/\sigma^2}{s}} (\mu_i - \mu_j) | \le q_{k,\nu}^{\alpha} \}$$

$$\times f_{\chi_\nu^2/\nu}(s)\, ds \tag{2.31}$$

$$\ge 1 - \int_{s=0}^{\infty} P\{ \max_{1 \le i,j \le k} | \frac{\xi_i - \xi_j}{\sqrt{s}} + \frac{\mu_i - \mu_j}{\sqrt{z}} | \le q_{k,\nu}^{\alpha} \} f_{\chi_\nu^2/\nu}(s)\, ds$$

$$= \beta_Q(\mu, n_0, z) \,,$$

where the fifth equation above follows from Lemma 2.3, while the inequality follows from the fact that

$$\sqrt{\frac{\tilde{N}/\sigma^2}{s}} = \sqrt{\frac{\tilde{N}/\sigma^2}{S^2/\sigma^2}} \ge \frac{1}{\sqrt{z}} \quad \text{by (2.30)} \,,$$

and the property (iii) of Lemma 2.1. So, the power function $\beta_{\tilde{Q}}(\mu, \sigma, n_0, z)$ clearly depends upon σ from equation (2.31). But a lower bound of $\beta_{\tilde{Q}}(\mu, \sigma, n_0, z)$ can be obtained as $\beta_Q(\mu, n_0, z)$ which is independent of σ. From equation (2.31) it is also easy to see that the following theorem is true.

Theorem 2.8

(i) Let $\tilde{\mu}_1 = (0, \ldots, 0, kb_0/(k-1)) \in \mathbf{R}^k$. Then for any real number $b_0 \ge 0$,

$$\inf_{\mu \in \tilde{\Theta}_{b_0}^1} \beta_{\tilde{Q}}(\mu, n_0, z) = \beta_{\tilde{Q}}(\tilde{\mu}_1, n_0, z).$$

(ii) Let $\tilde{\mu}_2 = (-b_0/2, 0, \ldots, 0, b_0/2) \in \mathbf{R}^k$. Then for any real number $b_0 \ge 0$,

$$\inf_{\mu \in \tilde{\Theta}_{b_0}^2} \beta_{\tilde{Q}}(\mu, n_0, z) = \beta_{\tilde{Q}}(\tilde{\mu}_2, n_0, z).$$

Thus, in order to guarantee the power of the two-stage sample Studentised range test for the difference zone $\tilde{\Theta}_{b_0}^i$ (i=1 or 2) to be no less than the prefixed number $\beta \in (\alpha, 1)$, we only need to guarantee $\beta_Q(\tilde{\mu}^i, n_0, z) \ge \beta$, by decreasing z for the selected n_0. Hochberg & Lachenbruch (1976) considered the problem of guaranteeing the power of the two-stage sample Studentised range test for the difference zone $\tilde{\Theta}^2_{b_0}$, but they failed to give a proper proof of the least favourable configuration. Some numerical examples can be seen in their paper.

Alternatively, the two-stage sample F-test operates as follows. Define

$$\tilde{L} = \frac{1}{(k-1)} \sum_{i=1}^{k} \Big(\frac{1}{\sqrt{\tilde{N}S^2}} \sum_{j=1}^{\tilde{N}} X_{ij} - \frac{1}{k\sqrt{\tilde{N}S^2}} \sum_{l=1}^{k} \sum_{j=1}^{\tilde{N}} X_{lj}\Big)^2 .$$

Then a size α test based on $\tilde{L}$ is to reject the null hypothesis H_0 iff

$$\tilde{L} \geq F^{\alpha}_{k-1,\nu} ,$$

where $F^{\alpha}_{k-1,\nu}$ is the upper α-point of the F distribution with parameters $k-1$ and $\nu = k(n_0-1)$. This test has size exactly equal to α since under the null hypothesis H_0 :

$$\tilde{L} = \frac{1}{(k-1)} \sum_{i=1}^{k} (\zeta_i - \bar{\zeta})^2 ,$$

which, as follows from Lemma 2.3, has a F distribution with parameters k and ν. The power of this test procedure, $\beta_{\tilde{L}}(\mu,\sigma,n_0,z)$ satisfies

$$\begin{aligned}
1 - \beta_{\tilde{L}}(\mu,\sigma,n_0,z) &= P\{\ \tilde{L} \leq F^{\alpha}_{k-1,\nu}\ \} \\
&= P\{\ \frac{1}{(k-1)} \sum_{i=1}^{k} \Big(\frac{1}{\sqrt{\tilde{N}S^2}} \sum_{j=1}^{\tilde{N}} X_{ij} - \frac{1}{k\sqrt{\tilde{N}S^2}} \sum_{l=1}^{k} \sum_{j=1}^{\tilde{N}} X_{lj}\Big)^2 \leq F^{\alpha}_{k-1,\nu}\ \} \\
&= P\{\ \frac{1}{(k-1)} \sum_{i=1}^{k} \Big[(\zeta_i + \mu_i\sqrt{\tilde{N}/S^2}) - \frac{1}{k}\sum_{l=1}^{k}(\zeta_l + \mu_l\sqrt{\tilde{N}/S^2})\Big]^2 \leq F^{\alpha}_{k-1,\nu}\ \} \\
&= P\{\ \frac{1}{(k-1)} \sum_{i=1}^{k} \Big[\Big(\frac{\xi_i}{\sqrt{S^2/\sigma^2}} + \mu_i\sqrt{\frac{\tilde{N}/\sigma^2}{S^2/\sigma^2}}\Big) - \\
&\qquad \frac{1}{k}\sum_{l=1}^{k}\Big(\frac{\xi_l}{\sqrt{S^2/\sigma^2}} + \mu_l\sqrt{\frac{\tilde{N}/\sigma^2}{S^2/\sigma^2}}\Big)\Big]^2 \leq F^{\alpha}_{k-1,\nu}\ \} \\
&= \int_{s=0}^{\infty} P\{\ \frac{1}{(k-1)} \sum_{i=1}^{k} \Big[\Big(\frac{\xi_i}{\sqrt{s}} + \mu_i\sqrt{\frac{\tilde{N}/\sigma^2}{s}}\Big) - \frac{1}{k}\sum_{l=1}^{k}\Big(\frac{\xi_l}{\sqrt{s}} + \mu_l\sqrt{\frac{\tilde{N}/\sigma^2}{s}}\Big)\Big]^2 \\
&\qquad \leq F^{\alpha}_{k-1,\nu}\ \} \times f_{\chi^2_\nu/\nu}(s)\ ds \qquad (2.32) \\
&\leq \int_{s=0}^{\infty} P\{\ \frac{1}{(k-1)} \sum_{i=1}^{k} \Big[\Big(\frac{\xi_i}{\sqrt{s}} + \frac{\mu_i}{\sqrt{z}}\Big) - \frac{1}{k}\sum_{l=1}^{k}\Big(\frac{\xi_l}{\sqrt{s}} + \frac{\mu_l}{\sqrt{z}}\Big)\Big]^2 \leq F^{\alpha}_{k-1,\nu}\ \} \\
&\qquad \times f_{\chi^2_\nu/\nu}(s)\ ds
\end{aligned}$$

$$= 1 - \beta_L(\mu, n_0, z) \ ,$$

where the fourth equation above follows from Lemma 2.3, while the inequality follows again from the fact that

$$\sqrt{\frac{\tilde{N}/\sigma^2}{s}} = \sqrt{\frac{\tilde{N}/\sigma^2}{S^2/\sigma^2}} \geq \frac{1}{\sqrt{z}} \quad \text{by (2.30)} \ ,$$

and the property that the right tail probability of the non-central χ^2 distribution increases as the non-centrality parameter increases. So, the power function $\beta_{\tilde{L}}(\mu,\sigma,n_0,z)$ also clearly depends upon σ from equation (2.32). But a lower bound of $\beta_{\tilde{L}}(\mu,\sigma,n_0,z)$ can be taken as $\beta_L(\mu,n_0,z)$ which is independent of σ. From equation (2.32) we have the following theorem.

Theorem 2.9

(i) Let $\tilde{\mu}_1 = (0,\ldots,0,kb_0/(k-1)) \in \mathbf{R}^k$. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu \in \tilde{\Theta}^1_{b_0}} \beta_{\tilde{L}}(\mu,\sigma,n_0,z) = \beta_{\tilde{L}}(\tilde{\mu}_1,\sigma,n_0,z).$$

(ii) Let $\tilde{\mu}_2 = (-b_0/2,0,\ldots,0,b_0/2) \in \mathbf{R}^k$. Then for any real number $b_0 \geq 0$,

$$\inf_{\mu \in \tilde{\Theta}^2_{b_0}} \beta_{\tilde{L}}(\mu,\sigma,n_0,z) = \beta_{\tilde{L}}(\tilde{\mu}_2,\sigma,n_0,z).$$

Thus, we can guarantee the power of the two-stage sample F-test for the difference zone $\tilde{\Theta}^i_{b_0}$ (i=1 or 2) to be no less than the preassigned number $\beta \in (\alpha, 1)$ by simply guaranteeing $\beta_L(\tilde{\mu}^i, n_0, z) \geq \beta$. This can be done by decreasing z for the selected n_0 following the calculation formulas (2.27), (2.28) and (2.29). We do not go into the details here.

Table 2.1

The least favourable power of the size $\alpha=0.05$ level Studentised range test for the difference zone

$$\Theta^1_{0.5} = \{ \mu : b_1(\mu) = \max_{1\le i\le k} |\mu_i-\bar{\mu}| \ge 0.5\sigma \}.$$

	k=3	4	5	6	7	8
$n = 5$	0.174	0.139	0.122	0.111	0.104	0.099
$n = 6$	0.209	0.165	0.143	0.129	0.120	0.113
$n = 8$	0.281	0.218	0.187	0.167	0.153	0.114
$n = 10$	0.351	0.274	0.233	0.208	0.191	0.178
$n = 15$	0.516	0.412	0.354	0.316	0.290	0.271
$n = 20$	0.654	0.540	0.472	0.426	0.393	0.369
$n = 25$	0.761	0.650	0.579	0.530	0.493	0.465
$n = 30$	0.840	0.740	0.672	0.623	0.586	0.556
$n = 35$	0.895	0.812	0.750	0.703	0.667	0.638
$n = 40$	0.933	0.866	0.812	0.771	0.737	0.709
$n = 45$	0.958	0.906	0.862	0.825	0.795	0.770
$n = 50$	0.974	0.935	0.899	0.869	0.843	0.821

Table 2.2

The least favourable power of the size $\alpha=0.05$ level Studentised range test for the difference zone

$$\Theta^1_{1.0} = \{ \mu : b_1(\mu) = \max_{1\le i\le k} |\mu_i-\bar{\mu}| \ge 1.0\sigma \}.$$

	k=3	4	5	6	7	8
$n = 5$	0.560	0.459	0.403	0.367	0.341	0.322
$n = 6$	0.666	0.559	0.496	0.454	0.424	0.401
$n = 8$	0.820	0.721	0.656	0.611	0.577	0.551
$n = 10$	0.909	0.833	0.778	0.736	0.704	0.678
$n = 15$	0.986	0.962	0.937	0.915	0.897	0.880
$n = 20$	0.998	0.993	0.985	0.977	0.970	0.963
$n = 25$	1.000	0.999	0.997	0.995	0.992	0.990
$n = 30$	1.000	1.000	0.999	0.999	0.998	0.997
$n = 35$	1.000	1.000	1.000	1.000	1.000	0.999
$n = 40$	1.000	1.000	1.000	1.000	1.000	1.000

Table 2.3

A comparison of the power functions of the size $\alpha=0.05$ level Studentised range test and F-test under their common least favourable configuration of population means for the difference zone

$$\Theta_{b_0}^1 = \{ \mu : b_1(\mu) = \max_{1 \le i \le k} | \mu_i - \bar{\mu} | \ge b_0 \sigma \}.$$

	b_0	power of the F-test	power of the SR test
$k=3$, $n=21$	0.31	0.308	0.304
	0.49	0.675	0.666
	0.62	0.866	0.857
$k=4$, $n=16$	0.43	0.338	0.336
	0.61	0.613	0.609
	0.78	0.845	0.840
$k=5$, $n=13$	0.55	0.368	0.373
	0.78	0.667	0.673
	1.00	0.891	0.892
$k=6$, $n=11$	0.67	0.397	0.411
	0.81	0.558	0.576
	1.08	0.839	0.852
$k=10$, $n=7$	1.13	0.496	0.564
	1.36	0.687	0.757
	1.59	0.841	0.892

Table 2.4

The least favourable power of the size $\alpha=0.05$ level Studentised range test for the difference zone

$$\Theta_{0.5}^2 = \{ \mu : b_2(\mu) = \max_{1\le i,j\le k} |\mu_i-\mu_j| \ge 0.5\sigma \}.$$

	k=3	4	5	6	7	8
$n = 5$	0.089	0.081	0.076	0.073	0.070	0.068
$n = 6$	0.099	0.089	0.083	0.079	0.075	0.073
$n = 8$	0.121	0.106	0.097	0.091	0.087	0.083
$n = 10$	0.143	0.124	0.112	0.104	0.099	0.094
$n = 15$	0.200	0.171	0.153	0.140	0.131	0.124
$n = 20$	0.260	0.221	0.197	0.179	0.166	0.157
$n = 25$	0.320	0.273	0.242	0.221	0.205	0.192
$n = 30$	0.379	0.325	0.290	0.265	0.246	0.231
$n = 35$	0.436	0.378	0.339	0.310	0.288	0.270
$n = 40$	0.491	0.429	0.387	0.356	0.331	0.312
$n = 45$	0.543	0.479	0.435	0.401	0.375	0.353
$n = 50$	0.592	0.527	0.481	0.446	0.418	0.396

Table 2.5

The least favourable power of the size $\alpha=0.05$ level Studentised range test for the difference zone

$$\Theta^2_{1.0} = \{ \mu : b_2(\mu) = \max_{1\le i,j\le k} |\mu_i - \mu_j| \ge 1.0\sigma \}.$$

	k=3	4	5	6	7	8
$n = 5$	0.221	0.190	0.171	0.158	0.149	0.141
$n = 6$	0.269	0.231	0.207	0.191	0.178	0.169
$n = 8$	0.365	0.316	0.283	0.261	0.243	0.229
$n = 10$	0.457	0.400	0.362	0.333	0.312	0.294
$n = 15$	0.654	0.592	0.547	0.512	0.485	0.462
$n = 20$	0.794	0.741	0.700	0.667	0.639	0.616
$n = 25$	0.883	0.844	0.812	0.785	0.761	0.740
$n = 30$	0.937	0.910	0.888	0.867	0.849	0.833
$n = 35$	0.967	0.951	0.936	0.922	0.909	0.897
$n = 40$	0.983	0.974	0.964	0.955	0.947	0.939
$n = 45$	0.992	0.986	0.981	0.975	0.970	0.965
$n = 50$	0.996	0.993	0.990	0.987	0.983	0.980

Table 2.6

A comparison of the power functions of the size $\alpha=0.05$ level Studentised range test and F-test under their common least favourable configuration of population means for the difference zone

$$\Theta^2_{b_0} = \{ \mu : b_2(\mu) = \max_{1\le i,j\le k} | \mu_i - \mu_j | \ge b_0\sigma \}.$$

	b_0	power of the F-test	power of the SR test
$k=3,\ n=21$	0.53 0.86 1.07	0.308 0.675 0.866	0.306 0.675 0.866
$k=4,\ n=16$	0.71 0.99 1.27	0.338 0.613 0.845	0.336 0.615 0.850
$k=5,\ n=13$	0.88 1.23 1.58	0.368 0.667 0.891	0.370 0.677 0.900
$k=6,\ n=11$	1.04 1.25 1.67	0.397 0.558 0.839	0.404 0.572 0.858
$k=10,\ n=7$	1.69 2.03 2.37	0.496 0.687 0.841	0.542 0.745 0.890

Chapter 3

The optimal tests for comparing normal means with known variance

---

3.1 Introduction

Consider the balanced one-way fixed effects analysis of variance model

$$Y_{ij} = \mu_i + \varepsilon_{ij} \qquad 1 \le i \le k, \ \ 1 \le j \le n$$

where the μ_i, $1 \le i \le k$, are k unknown treatment means and the ε_{ij} are i.i.d. $N(0, \sigma^2)$ random variables with known variance σ^2. Suppose that an experimenter intends initially to perform a size α hypothesis test of the null hypothesis that the treatment means are all equal, $H_0 : \mu_1 = \ldots = \mu_k$, against the general alternative hypothesis "not H_0", and to assess the power of such a test. As was stated in chapter 2, we could employ two test procedures: one test uses the sum of squares statistic, $\sum_{i=1}^{k} (X_i - \bar{X})^2$, with $X_i = \bar{Y}_i$, $1 \le i \le k$, and the other test uses the range statistic, $\max_{1 \le i, j \le k} |X_i - X_j|$. If the sum of squares test statistic, $\sum_{i=1}^{k} (X_i - \bar{X})^2$, is used, then the power of the test procedure depends on the treatment means only through the quantity $\varepsilon^2 = \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2$. It is generally agreed, however, that assessing the power levels in terms of ε^2 is not easily interpretable by the experimenter, and that it is more useful to assess power levels in terms of certain range measures such as $b_1(\mu) = \max_{1 \le i \le k} |\mu_i - \bar{\mu}|$ or $b_2(\mu) = \max_{1 \le i,j \le k} |\mu_i - \mu_j|$ (see, for example, Pearson & Hartley (1951), and Scheffe (1959)p.62). If the range test statistic, $\max_{1 \le i,j \le k} |X_i - X_j|$ is used, we have the results in chapter 2 about the least favourable configurations of the power function of this test for the difference zones determined by $b_1(\mu)$ and $b_2(\mu)$. For the purpose of testing the null hypothesis H_0, it also seems reasonable to use the test statistic $\max_{1 \le i \le k} |X_i - \bar{X}|$ where $\bar{X} = \frac{1}{k}\sum_{i=1}^{k} X_i$. The question now to be answered is: amongst all suitable test procedures of size α, which test maximises the minimum power in the region of the parameter space defined by $b_i(\mu) \ge \delta_0$ (i=1 or 2), for a fixed value of δ_0? In other words, we want to find the maximin test procedures for the difference zones

$$\tilde{\Theta}^1_{\delta_0} = \{ \mu : \max_{1 \le i \le k} |\mu_i - \bar{\mu}| \ge \delta_0 \} \tag{3.1}$$

and

$$\tilde{\Theta}^2_{\delta_0} = \{ \mu : \max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge \delta_0 \} . \tag{3.2}$$

The derivation of such optimal test procedures is important for two reasons. Firstly, in terms of experimental design, an experimenter may decide to choose the sample size n so that for specified values of α, β and δ, a size α test will have power of at least β, whenever $\max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge \delta_0$ say, then the use of the optimal test procedure will allow the probability requirements to be met most economically with the smallest possible sample size. Secondly, comparing the optimal test procedures with the three commonly used multiple comparison test procedures: the F-test, the Studentised range test and the comparisons with average test procedure (see, for example, Hochberg & Tamhane (1987)) when the variance σ^2 is unknown, sheds interesting light on the power properties of these common test procedures.

We now define a class Ψ of tests $\psi(\mathbf{x})$ (we confine ourselves to non-random tests) in which the search for the maximin test procedure will be carried out. A test is a function $\psi : \mathbf{R}^k \to \{0,1\}$, where we suppose that $\psi(\mathbf{x}) = 1$ implies that the null hypothesis H_0 is rejected, and $\psi(\mathbf{x}) = 0$ implies that H_0 is accepted. A test $\psi(\mathbf{x})$ is symmetric iff

$$\psi(\, \mathbf{x} \,) = \psi(\, -\mathbf{x} \,) \quad \text{for all } \mathbf{x} \in \mathbf{R}^k ,$$

and exchangeable iff

$$\psi(\,\pi\mathbf{x}\,) \;=\; \psi(\,\mathbf{x}\,) \qquad \text{for all } \mathbf{x} \in \mathbf{R}^k,\;\; \pi \in \Pi,$$

where Π is the set of all $k!$ permutation transformations of the coordinates. Furthermore, we call a test shift invariant iff

$$\psi(\,\mathbf{x} + c\mathbf{1}\,) = \psi(\,\mathbf{x}\,)\,, \qquad \text{for all } \mathbf{x} \in \mathbf{R}^k,\;\; c \in \mathbf{R}$$

where $\mathbf{1} = (1,1,\ldots,1)$, and convex iff

$$\psi(\,\mathbf{x}\,) = \psi(\,\mathbf{y}\,) = 0 \;\Rightarrow\; \psi(\,\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\,) = 0\,, \qquad \text{for all } \lambda \in (0,1),\;\; \mathbf{x},\,\mathbf{y} \in \mathbf{R}^k.$$

The class Ψ of tests $\psi(\mathbf{x})$ is then defined by

$$\Psi = \{\; \psi(\mathbf{x}) : \;\psi \text{ a symmetric, exchangeable, shift invariant, convex test } \}\,.$$

A maximin test procedure is now taken to mean the maximin test procedure within the class Ψ. It is clear that any sensible test of the null hypothesis H_0 will be in the class Ψ. The properties of symmetry, permutation invariance and shift invariance are clearly desirable, as is the property of convexity since $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$, $\lambda \in (0,1)$ may be thought of as a "smoothed" version of the data $\mathbf{x}$ and $\mathbf{y}$, which indicates more strongly the possibility of the null hypothesis H_0 holding than does either $\mathbf{x}$ or $\mathbf{y}$ alone. The three tests mentioned above are clearly all in the class Ψ, as are the tests using the following sets as critical regions,

$$\{\,\mathbf{X} : \sum_{1\le i,j\le k} |X_i - X_j|^l \ge c\,\}, \qquad l \ge 1$$

$$\{\,\mathbf{X} : \sum_{i=1}^{k} |X_i - \bar{X}|^l \ge c\,\}, \qquad l \ge 1$$

or more generally

$$\{\,\mathbf{X} : \sum_{1\le i,j\le k} g(\,|X_i - X_j|\,) \ge c\,\},$$

$$\{\,\mathbf{X} : \sum_{i=1}^{k} g(\,|X_i - \bar{X}|\,) \ge c\,\},$$

where the function $g : [0,\infty) \to \mathbf{R}$ is non-negative, increasing and convex.

Next, we introduce some notation for defining difference zones. Let

$$\mathbf{c}_1(j) = (\,(j-k)/k,\,\ldots\,,(j-k)/k,\; j/k,\,\ldots\,,\,j/k\,), \quad 1 \le j \le k/2,$$

where the first j terms are identical, as are the last $k-j$ terms, and let

$$\mathbf{c}_2(j) = (\ -1,\ \dots\ ,\ -1,\ 0,\ \dots\ ,\ 0,\ 1,\ \dots\ ,\ 1\ ), \quad 1 \le j \le k/2,$$

where there are j terms each of -1 and 1. Also define the difference zone $\Theta_A(\mathbf{c},\delta)$ to be

$$\Theta_A(\mathbf{c},\delta) = \{\ \boldsymbol{\mu} :\ \max_{\pi\in\Pi} |\ (\pi\mathbf{c})'\boldsymbol{\mu}\ | \ge \delta\ \}.$$

Then $\tilde{\Theta}^1_{\delta_0} = \Theta_A(\mathbf{c}_1(1),\delta_0)$ and $\tilde{\Theta}^2_{\delta_0} = \Theta_A(\mathbf{c}_2(1),\delta_0)$, where $\tilde{\Theta}^i_{\delta_0}$, $i=1,\ 2$ are defined in (3.1) and (3.2). In section 3.3, the maximin tests within the class Ψ will be found for slightly more general settings than discussed above where $\mathbf{X} \sim N_k(\boldsymbol{\mu},\ \Sigma)$ for the matrix Σ which has diagonal elements all equal σ^2, and equal off-diagonal elements $\rho\sigma^2$, and where the difference zones are $\Theta_A(\mathbf{c}_i(j),\delta)$, for $i=1$ or 2 and any j, $1\le j\le k/2$.

The derivation of maximin tests in section 3.3 depends upon the results of section 3.2 where various least favourable configurations of $\boldsymbol{\mu}$ are found for the class of tests Ψ. In fact, the derivations in section 3.2 are carried out under a weaker distribution condition than the normality of the sample $\mathbf{X} = (\ X_1,\ \dots\ ,\ X_k\ )$. We only assume that the probability density function of $\mathbf{X}$ is $f(\mathbf{x}-\boldsymbol{\mu})$ where $f(\mathbf{x})$ is a member of the class of densities $\mathbf{F}$ defined by

$$\mathbf{F} = \{\ f(\mathbf{x}) :\ f(\mathbf{x}) \text{ is symmetric, exchangeable, unimodal } \}.$$

In this case unimodal is taken to mean that the set $\{\ \mathbf{x} :\ f(\mathbf{x}) \ge c\ \}$ is a convex set for all $c\in\mathbf{R}$. So, $f(\mathbf{x}) = \varphi(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}\}$, where Σ is as above, is clearly in F. In the case where X_i, $1\le i\le k$ are independent random variables with probability density $f_0(x-\mu_i)$, we have, following a result of Mudholkar (1969) (see Theorem A3 in the appendix), that if $f_0(x)$ is symmetric and log-concave (see appendix for this definition) then $f(\mathbf{x}) = \prod_{i=1}^{k} f_0(x_i)$ is a member of $\mathbf{F}$.

It is easy to check that the following univariate density functions, with the location parameter set to zero, are symmetric and log-concave

$$\text{Normal} \quad f_0(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{x^2}{2\sigma^2}\}, \quad x \in (-\infty,\ \infty),\ \sigma^2 > 0\ .$$

$$\text{Uniform} \quad f_0(x) = \frac{1}{2a}I_{[-a,\ a]}(x), \quad a > 0\ .$$

$$\text{Laplace} \quad f_0(x) = \frac{1}{2\beta}\exp\{-\frac{|x|}{\beta}\}, \quad x \in (-\infty,\ \infty),\ \beta > 0\ .$$

Triangular $\quad f_0(x) = \frac{1}{a}(1-\frac{|x|}{a})I_{[-a,\, a]}(x), \quad a > 0 .$

The results of section 3.2 are important not only because they are necessary for the derivation of the maximin test procedures in section 3.3 but also in their own right. In the situations where the probability density of $\mathbf{X}$ is $f(\mathbf{x}-\mu)$ with $f(\mathbf{x})$ in $\mathbf{F}$, and the test procedure employed for testing the null hypothesis H_0 is in Ψ, an experimenter can guarantee the power requirements for the difference zones $\Theta_A(\mathbf{c}_i(j),\delta)$ by guaranteeing the power values at the least favourable configuration derived explicitly in section 3.2.

3.2 Derivation of the least favourable configurations

In this section we discuss the identification of the least favourable configurations of the parameters $\mu \in \mathbf{R}^k$, which are defined to be those μ which minimize the power of the test procedure within the difference zone under consideration. For difference zones of the form $\Theta_A(\mathbf{c}_i(j),\delta)$, we shall explicitly find a least favourable configuration common to all densities $f(\mathbf{x}) \in \mathbf{F}$ and to all tests $\psi(\mathbf{x}) \in \Psi$.

For a given test procedure $\psi(\mathbf{x})$, let $A \subset \mathbf{R}^k$ be the respective acceptance region, and denote the power function of the test ψ, for a given set of parameters μ and density $f(\mathbf{x})$, by $\beta(\psi,f,\mu)$, so that

$$\beta(\psi,f,\mu) = 1 - \int_A f(\mathbf{x}-\mu)d\mathbf{x}. \tag{3.3}$$

Notice that $\psi(\mathbf{x}) \in \Psi$ implies that the acceptance region A is symmetric about the origin, exchangeable, convex and shift invariant.

The following lemma establishes some basic properties of the power function $\beta(\psi,f,\mu)$.

Lemma 3.1 Let $f(\mathbf{x}) \in \mathbf{F}$, $\psi(\mathbf{x}) \in \Psi$ and $\mu \in \mathbf{R}^k$. Then we have

(i) $\quad \beta(\psi,f,\mu) = \beta(\psi,f,\dot{-}\mu)$.

(ii) $\quad \beta(\psi,f,\mu) = \beta(\psi,f,\pi\mu), \quad$ for all $\pi \in \Pi$.

(iii) $\quad \beta(\psi,f,\mu) = \beta(\psi,f,\mu+c\mathbf{1}), \quad$ for all $c \in \mathbf{R}$.

(iv) $\beta(\psi,f,\mu) \geq \beta(\psi,f,\lambda\mu)$, $\lambda \leq 1$.

(v) $\beta(\psi,f,\mu) \geq \beta(\psi,f,\xi)$, if $\mu \gg \xi$.

Proof of Lemma 3.1

Properties (i)-(iii) of the lemma follow immediate from the definitions of the classes Ψ and **F** and the representation (3.3). Property (iv) is a consequence of a result of Anderson (1955) (Theorem A4 in the appendix). Property (v), where $\mu \gg \xi$ means μ majorises ξ, is a consequence of a result of Marshall & Olkin (1974) (Theorem A1 in the appendix) #

From properties (i), (ii) and (iii) of Lemma 3.1, we can see that a least favourable configuration will remain so after the operations in (i), (ii) and (iii) or combinations thereof, provided that the mean vector after these operations is still in the difference zone under consideration. Notice that any μ in the difference zone $\Theta_A(\mathbf{c}_i(j),\delta)$ will remain in it after any combination of the above three operations. So, if we find a single least favourable configuration for $\Theta_A(\mathbf{c}_i(j),\delta)$, we can then find a set of least favourable configurations.

Now we start the results of this section. First, the following theorem establishes a least favourable configuration for the difference zone $\Theta_A(\mathbf{c}_1(j),\delta)$ common to all densities $f(\mathbf{x}) \in \mathbf{F}$ and test $\psi(\mathbf{x}) \in \Psi$.

Theorem 3.1 Let $f(\mathbf{x}) \in \mathbf{F}$, $\psi(\mathbf{x}) \in \Psi$, and $\delta > 0$. Define

$$\mu_1(j) = (-\delta/j, \dots , -\delta/j, \delta/(k-j), \dots , \delta/(k-j)) \tag{3.4}$$

where $-\delta/j$ occupies the first j places. Then for any j, $1 \leq j \leq k/2$, we have

$$\beta(\psi,f,\mu_1(j)) \leq \beta(\psi,f,\mu)$$

for any $\mu \in \Theta_A(\mathbf{c}_1(j),\delta)$.

Proof of Theorem 3.1

As a consequence of property (iv) of Lemma 3.1, it is sufficient to restrict attention to the vectors μ for which $\max\limits_{\pi\in\Pi} |(\pi\mathbf{c}_1(j))'\mu| = \delta$. Also, from the properties (ii) and (iii) of Lemma 3.1 we can assume $\mu_1 \leq \mu_2 \leq \dots \leq \mu_k$ and $\sum\limits_{i=1}^{k}\mu_i = 0$ without loss of generality. Then as a consequence of a result of Harty (1952) (see Theorem A5 in appendix) we have

$$\delta = \max\{ \ |\sum_{i=1}^{k} c_1(j)_i \ \mu_i|, \ |\sum_{i=1}^{k} c_1(j)_i \ \mu_{k+1-i}| \ \}$$

and again, without loss of generality we can assume $\delta = |\sum_{i=1}^{k} c_1(j)_i \ \mu_i|$ (otherwise consider $\mu' = (-\mu_k, \dots , -\mu_1)$ instead of μ). Now, using the assumptions that $\mu_1 \le \mu_2 \le \dots \le \mu_k$ and $\sum_{i=1}^{k}\mu_i = 0$, we have $\delta = \sum_{i=1}^{k} c_1(j)_i \mu_i$. Thus,

$$\delta = \frac{(j-k)}{k}\sum_{i=1}^{j}\mu_i + \frac{j}{k}\sum_{i=j+1}^{k}\mu_i, \quad \text{and} \quad \sum_{i=1}^{k}\mu_i = 0$$

which imply that $\sum_{i=1}^{j}\mu_i = -\delta$, $\sum_{i=j+1}^{k}\mu_i = \delta$. This, together with the condition $\mu_1 \le \dots \le \mu_k$ implies

$$\begin{aligned} \mu &= (\mu_1, \dots , \mu_j, \mu_{j+1}, \dots , \mu_k) \\ &\gg (\sum_{i=1}^{j}\mu_i/j, \dots , \sum_{i=1}^{j}\mu_i/j, \sum_{i=j+1}^{k}\mu_i/(k-j), \dots , \sum_{i=j+1}^{k}\mu_i/(k-j)) \\ &= (-\delta/j, \dots , -\delta/j, \delta/(k-j), \dots , \delta/(k-j)) = \mu_1(j). \end{aligned}$$

By appealing to property (v) of Lemma 3.1, the proof is therefore completed #

From Theorem 3.1 and the fact that $\Theta_A(c_1(1),\delta) = \{ \mu: \max_{1\le i\le k}|\mu_i-\bar{\mu}| \ge \delta \}$, we immediately have the following corollary which includes Theorem 2.2 and Theorem 2.3 of chapter 2 as special cases.

Corollary 3.1 Let $f(\mathbf{x}) \in \mathbf{F}$, $\psi(\mathbf{x}) \in \Psi$ and $\delta > 0$, then a least favourable configuration of test $\psi(\mathbf{x})$ for the difference zone $\{ \mu : \max_{1\le i\le k}|\mu_i-\bar{\mu}| \ge \delta \}$ is $\mu_1(1) = (-\delta, \delta/(k-1), \dots , \delta/(k-1))$.

From Theorem 3.1 we can also see that the usual one direction slippage alternative hypothesis is the least favourable configuration of the difference zone $\Theta_A(c_1(j),\delta)$ for tests in Ψ. Now we turn our attention to the difference zone $\Theta_A(c_2(j),\delta)$. The following theorem establishes a least favourable configuration for the difference zone $\Theta_A(c_2(j),\delta)$ which is common to all densities $f(\mathbf{x}) \in \mathbf{F}$ and tests $\psi(\mathbf{x}) \in \Psi$.

Theorem 3.2 Let $f(\mathbf{x}) \in \mathbf{F}$, $\psi(\mathbf{x}) \in \Psi$ and $\delta > 0$. Define

$$\mu_2(j) = (-\delta/(2j), \dots , -\delta/(2j), 0,\dots,0, \delta/(2j), \dots , \delta/(2j)), \qquad (3.5)$$

where $-\delta/(2j)$ and $\delta/(2j)$ each occupy j places. Then for any j, $1 \le j \le k/2$,

$$\beta(\psi, f, \mu_2(j)) \le \beta(\psi, f, \mu)$$

for any $\mu \in \Theta_A(\mathbf{c}_2(j), \delta)$.

Proof of Theorem3.2

By property (iv) of Lemma 3.1, we only need to consider the vectors μ which satisfy $\max_{\pi \in \Pi} |(\pi \mathbf{c}_2(j))'\mu| = \delta$. By the properties (ii) and (iii) of Lemma 3.1, we can assume $\mu_1 \le \mu_2 \le \dots \le \mu_k$ and $\sum_{i=1}^{j} \mu_i = -\sum_{i=k-j+1}^{k} \mu_i$ without loss of generality. Then following Theorem A5 in the appendix, we have

$$\delta = \max\{ \, |\sum_{i=1}^{k} \mathbf{c}_2(j)_i \, \mu_i|, \ |\sum_{i=1}^{k} \mathbf{c}_2(j)_i \, \mu_{k+1-i}| \, \}$$

and again, without loss of generality, we can assume $\delta = |\sum_{i=1}^{k} \mathbf{c}_2(j)_i \, \mu_i|$ (otherwise consider $\mu' = (-\mu_k, \dots , -\mu_1)$ instead of μ). Now, using the assumption that $\mu_1 \le \mu_2 \le \dots \le \mu_k$ and the definition of $\mathbf{c}_2(j)$, we have $\delta = \sum_{i=1}^{k} \mathbf{c}_2(j)_i \mu_i$. Thus we have

$$\delta = -\sum_{i=1}^{j} \mu_i + \sum_{i=k-j+1}^{k} \mu_i = -2 \sum_{i=1}^{j} \mu_i = 2 \sum_{i=k-j+1}^{k} \mu_i \, .$$

So

$$\begin{aligned} \mu &= (\mu_1, \dots , \mu_j, \mu_{j+1}, \dots , \mu_{k-j}, \mu_{k-j+1}, \dots , \mu_k) \\ &\gg (\sum_{i=1}^{j} \mu_i/j, \dots , \sum_{i=1}^{j} \mu_i/j, \sum_{i=j+1}^{k-j} \mu_i/(k-2j), \dots , \sum_{i=j+1}^{k-j} \mu_i/(k-2j), \\ &\qquad \sum_{i=k-j+1}^{k} \mu_i/j, \dots , \sum_{k-j+1}^{k} \mu_i/j) \\ &= (-\delta/2j, \dots , -\delta/2j, a, \dots , a, \delta/2j, \dots , \delta/2j) \equiv \mu_2(a,j), \end{aligned}$$

where $a = \sum_{i=j+1}^{k-j} \mu_i/(k-2j)$, $|a| \le -\delta/2$, and $-\delta/(2j)$ and $\delta/(2j)$ each occupy j places. Appealing to property (v) of Lemma 3.1, we have

$$\beta(\,\psi,f,\mu\,) \geq \beta(\,\psi,f,\mu_2(a,j)\,).$$

So, the proof will be completed if we can establish that

$$F(a) = 1-\beta(\,\psi,f,\mu_2(a,j)\,) = \int_{\mathbf{A}} f(\mathbf{x}-\mu_2(a,j))\;dx$$

is a decreasing function of $|a|$. To see this, consider the change of variables

$$y_i = (x_i - x_{k+1-i})/2 \qquad 1 \leq i \leq j$$

$$z_i = (x_i + x_{k+1-i})/2 \qquad 1 \leq i \leq j$$

$$w_i = x_{j+i} \qquad 1 \leq i \leq k-2j,$$

then we have

$$F(a) = \int_{A_k(\mathbf{y}\ \mathbf{w}\ \mathbf{z})} g(\,\mathbf{y},\ \mathbf{w}-a\mathbf{1},\ \mathbf{z}\,)\ d\mathbf{y}d\mathbf{w}d\mathbf{z}$$

where

$$g(\,\mathbf{y},\ \mathbf{w},\ \mathbf{z}) = 2^j\ f(\ z_1+y_1+\delta/2j,\ \dots\ ,\ z_j+y_j+\delta/2j,\ w_1,\ \dots\ ,\ w_{k-2j},$$
$$z_j - y_j - \delta/2j,\ \dots\ ,\ z_1 - y_1 - \delta/2j\ )$$

and

$$A_k(\mathbf{y}\ \mathbf{w}\ \mathbf{z}) = \{\ (\mathbf{y},\mathbf{w},\mathbf{z}) :\ \mathbf{y} \in \mathbf{R}^j,\ \mathbf{w} \in \mathbf{R}^{k-2j},\ \mathbf{z} \in \mathbf{R}^j,\ \mathbf{x} \in A\ \}.$$

Now, if we define the set $\tilde{A}(\mathbf{y}) \subset \mathbf{R}^{k-j}$ by

$$\tilde{A}(\mathbf{y}) = \{\ (\mathbf{z},\ \mathbf{w}) :\ \mathbf{z} \in \mathbf{R}^j,\ \mathbf{w} \in \mathbf{R}^{k-2j},\ \mathbf{x} \in A\ \},$$

then,

$$F(a) = \int_{\mathbf{y}\in\mathbf{R}^j} G(\ a,\ \mathbf{y}\ )\ d\mathbf{y} \tag{3.6}$$

with

$$G(\ a,\ \mathbf{y}\ ) = \int_{\tilde{A}(\mathbf{y})} g(\,\mathbf{y},\ \mathbf{w}-a\mathbf{1},\ \mathbf{z}\,)\ d\mathbf{w}d\mathbf{z}\ .$$

Now, we want to show that $G(\ a,\ \mathbf{y}\ )$ is a decreasing function of $|a|$ for fixed $\mathbf{y} \in \mathbf{R}^j$. Notice the conditions that the set A is symmetric about the origin, exchangeable and convex are sufficient to ensure that the set $\tilde{A}(\mathbf{y})$ is symmetric about the origin and convex. In addition, the conditions that the density $f(\mathbf{x})$ is symmetric about the origin, exchangeable and convex are sufficient to ensure that for a fixed value $\mathbf{y}$, $g(\mathbf{y},\mathbf{z},\mathbf{w})$ is, as a function of $(\mathbf{z},\mathbf{w})$, symmetric about the origin and unimodal. Therefore, Theorem A4 in the appendix may be applied to show that $G(a,\mathbf{y})$ is a decreasing function of $|a|$ for any fixed $\mathbf{y}$, and so that the required properties of $F(a)$ follow from equation (3.6). This completes the proof of Theorem 3.2 #

From Theorem 3.2 and the fact that $\Theta_A(\mathbf{c}_2(1),\delta) = \{ \mu : \max_{1\le i,j\le k} |\mu_i - \mu_j| \ge \delta \}$, we immediately have the following corollary which includes Theorem 2.4 and Theorem 2.5 of chapter 2 as special cases.

Corollary 3.2 Let $f(\mathbf{x}) \in \mathbf{F}$, $\psi(\mathbf{x}) \in \Psi$ and $\delta > 0$, then a least favourable configuration of test $\psi(\mathbf{x})$ for the difference zone $\{ \mu : \max_{1\le i,j\le k} |\mu_i - \mu_j| \ge \delta \}$ is $\mu_1(1) = (-\delta/2, 0, \dots , 0, \delta/2)$.

The results of this section are useful by themselves, since for a given test procedure, $\psi(\mathbf{x})$, and density, $f(\mathbf{x})$, an experimenter may assess the power level within a difference zone by evaluating the power at the least favourable configuration. Furthermore, the fact that the least favourable configurations are identical for this large family of tests allows the simple derivation of the maximin test procedures in the next section.

Before beginning the next section, we prove in addition a theorem concerning the least favourable configuration for tests in Ψ, and the multivariate normal density with a special form of covariance matrix.

Theorem 3.3 Let $f_0(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}'V^{-1}\mathbf{x}\}$, where the matrix $V = (v_{ij})_{k\times k}$ satisfies $v_{ii}+v_{jj}-2v_{ij} = 2d > 0$, $1 \le i \ne j \le k$. Then for $\psi(\mathbf{x}) \in \Psi$, $\delta>0$ and $1 \le j \le k/2$,

(i) $\beta(\psi,f_0,\mu_1(j)) \le \beta(\psi,f_0,\mu)$, for any $\mu \in \Theta_A(\mathbf{c}_1(j),\delta)$.

(ii) $\beta(\psi,f_0,\mu_2(j)) \le \beta(\psi,f_0,\mu)$, for any $\mu \in \Theta_A(\mathbf{c}_2(j),\delta)$.

Proof of Theorem 3.3

Let $A^c \subset \mathbf{R}^k$ be the critical region of the test $\psi(\mathbf{x})$. Notice that $\psi(\mathbf{x}) \in \Psi$ implies that A^c is shift invariant, so that

$$\beta(\psi,f_0,\mu) = P\{ \mathbf{X} \in A^c \}$$

$$= P\{ \mathbf{X} - X_k\mathbf{1} \in A^c \}$$

$$= P\{\ (X_1 - X_k,\ \dots\ ,\ X_{k-1} - X_k,\ 0) \in A^c\ \}\ , \tag{3.7}$$

where

$$\mathbf{X} - \boldsymbol{\mu} \sim f_0(\mathbf{x})\ .$$

Thus, $\beta(\psi, f_0, \boldsymbol{\mu})$ depends on the distribution of $\mathbf{X}$ only through the distribution of $(X_1 - X_k, \dots, X_{k-1} - X_k)$, and in this case

$$(\ X_1 - X_k,\ \dots\ ,\ X_{k-1} - X_k\ ) \sim N_{k-1}\ (\ \boldsymbol{\mu}_{k-1},\ V_{k-1}\ )\ ,$$

where $\boldsymbol{\mu}_{k-1} = (\mu_1 - \mu_k, \dots, \mu_{k-1} - \mu_k)$, and V_{k-1} is a $(k-1)\times(k-1)$ matrix with diagonal elements all equal to $2d$, off-diagonal elements all equal to d. Now, if $f_0^*(\mathbf{x}) \propto \exp\{-1/(2d)\mathbf{x}'\mathbf{x}\}$, we can see from equation (3.7) that

$$\beta(\psi, f_0, \boldsymbol{\mu}) = \beta(\psi, f_0^*, \boldsymbol{\mu}), \quad \text{for any } \boldsymbol{\mu}\ . \tag{3.8}$$

Since $f_0^*(\mathbf{x}) \in \mathbf{F}$, (i) and (ii) follow immediately from equation (3.8), and Theorem 3.1 and Theorem 3.2 respectively #

3.3 The maximin test procedures

In this section we derive maximin tests for difference zones of the form $\Theta_A(\mathbf{c}_1(j), \delta)$ and $\Theta_A(\mathbf{c}_2(j), \delta)$ when $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$, where Σ is a $k \times k$ matrix with diagonal elements all equal to σ^2, off-diagonal elements all equal to $\rho\sigma^2$, and the condition $-1/(k-1) < \rho < 1$ is satisfied (which is necessary and sufficient for Σ to be positive-definite, and hence a non-singular covariance matrix). Thus,

$$f(\mathbf{x}) = \varphi(\mathbf{x}) \propto exp\ \{-\tfrac{1}{2}\mathbf{x}'\Sigma^{-1}\mathbf{x}\}.$$

Theorem 3.4 below gives the maximin test for an difference zone of the form $\Theta_A(\mathbf{c}_1(j), \delta)$.

Theorem 3.4 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ which have size $\alpha \in (0, 1)$, the maximin test $\psi^*(\mathbf{x})$ for the difference zone $\Theta_A(\mathbf{c}_1(j), \delta)$ is given by

$$\psi^*(\mathbf{x}) = 1 \iff \sum_J \cosh\ \{w(\bar{x}_J - \bar{x})\} \geq c \tag{3.9}$$

where the sum is over all subsets J of $\{1,\ldots,k\}$ of size j, $\bar{x}_J$ is the average of the elements of $\mathbf{x}$ with indices in J, $\bar{x}$ is the average of all the elements of $\mathbf{x}$,

$$w = \frac{k\delta}{\sigma^2(k-j)(1-\rho)}$$

and the constant c is chosen so that the test has size α.

Proof of Theorem3.4

As a consequence of Theorem 3.1, it is apparent that the maximin test $\psi^*(\mathbf{x})$ is the test which maximizes

$$L(\psi) = \int_{\mathbf{R}^k} \psi(\mathbf{x})\ \varphi(\ \mathbf{x}-\mu_1(j)\ )\ d\mathbf{x}$$

over all tests $\psi(\mathbf{x}) \in \Psi$ subject to the size condition,

$$\alpha = \int_{\mathbf{R}^k} \psi(\mathbf{x})\ \varphi(\mathbf{x})\ d\mathbf{x}.$$

Now, since the density $\varphi(\mathbf{x})$ is symmetric about the origin, and any test $\psi(\mathbf{x}) \in \Psi$ is also symmetric about the origin and exchangeable, we have

$$L(\psi) = \frac{1}{2\ k!} \int_{\mathbf{R}^k} \psi(\mathbf{x})\ G(\mathbf{x})\ dx$$

where

$$\begin{aligned} G(\mathbf{x}) &= \sum_{\pi\in\Pi} \left[\varphi(\pi\mathbf{x}-\mu_1(j)) + \varphi(-\pi\mathbf{x}-\mu_1(j))\right] \\ &= \sum_{\pi\in\Pi} \left[\varphi(\pi\mathbf{x}-\mu_1(j)) + \varphi(\pi\mathbf{x}+\mu_1(j))\right]. \end{aligned}$$

Then, as a consequence of the Neyman-Pearson Lemma, the test given by

$$\psi^*(\mathbf{x}) = 1 \iff \frac{G(\mathbf{x})}{\varphi(\mathbf{x})} \geq c' \tag{3.10}$$

where the constant c' is chosen to satisfy the size requirement, will be the maximin test provided that it is in the class Ψ. However,

$$\frac{G(\mathbf{x})}{\varphi(\mathbf{x})} \propto \sum_{\pi\in\Pi} \cosh\{\ (\pi\mathbf{x})'\Sigma^{-1}\mu_1(j)\ \} \tag{3.11}$$

and so it is readily seen that the test given by equation (3.10) is symmetric, exchangeable, and convex since $\cosh(x)$ is a convex function. Also the test is shift invariant since $\mathbf{1}'\ \Sigma^{-1}\ \mu_1(j) = 0$, which follows from the fact that

$$\mathbf{1}'\,\Sigma^{-1} \propto \mathbf{1}' \quad \text{and} \quad \mathbf{1}'\,\mu_1(j) = 0\ ,$$

where $\mathbf{1} = (1,\ldots,1) \in \mathbf{R}^k$. Hence the test given in equation (3.10) is the maximin test, and it remains only to show that it is equivalent to the test given in equation (3.9). This is clear from the fact that

$$\Sigma^{-1}\,\mu_1(j) = \frac{1}{\sigma^2(1-\rho)}\,\mu_1(j)$$

and from equation (3.11), and so the proof of Theorem 3.4 is completed #

Notice that the maximin test given in Theorem 3.4 depends on δ, the variability of the difference zone, but depends on the size α only through the critical point. For the difference zone given in (3.1), it is clear that we have the following corollary.

Corollary 3.3 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ which have size $\alpha \in (0, 1)$, the maximin test $\psi^*(\mathbf{x})$ for the difference zone $\{\ \boldsymbol{\mu}\!: \max_{1\le i\le k}|\mu_i - \bar{\mu}| \ge \delta\ \}$ is given by

$$\psi^*(\mathbf{x}) = 1 \quad \Longleftrightarrow \quad \sum_{i=1}^{k} \cosh\{w'(x_i - \bar{x})\} \ge c \tag{3.12}$$

where $w' = \dfrac{k\,\delta}{\sigma^2(k-1)(1-\rho)}$, and the constant c is chosen to give the required size.

In Corollary 3.4 below we consider what happens to the maximin test as the value of δ approaches zero or approaches infinity. As δ approaches zero, we see that the maximin test becomes equivalent to a test based on the sum of squares statistic. The two tests are equivalent in the sense that for almost any data $\mathbf{x}$, the two tests make the same decision for small enough values of δ. At the other extreme, as δ approaches infinity, the maximin test becomes equivalent to a test based on a range statistic such as $\max_{1\le i\le k}|x_i - \bar{x}|$. To prove Corollary 3.4 and Corollary 3.6 below, we need the following lemma first.

Lemma 3.2 For real numbers $a_1, a_2, \ldots, a_k$, we have

(i) $$\sum_{(i_1,\ldots,i_m)} \Big(\sum_{r=1}^{m} a_{i_r} - m\bar{a}\Big)^2 = \frac{m(k-m)C_m^k}{k(k-1)} \sum_{i=1}^{k}(a_i - \bar{a})^2\ , \qquad \text{for } 1 \le m \le k.$$

(ii) $$\sum_{\substack{(i_1,\ldots,i_m)\\(j_1,\ldots,j_m)}} \left(\sum_{r=1}^{m} a_{i_r} - \sum_{s=1}^{m} a_{j_s} \right)^2 = \frac{2mC_m^k C_m^{k-m}}{k-1} \sum_{i=1}^{k} (a_i - \bar{a})^2 , \quad \text{for } 1 \le m \le k/2.$$

where $\bar{a} = \sum_{i=1}^{k} a_i / k$, the sum in (i) is over all subsets $(i_1, \ldots, i_m)$ of $\{1, \ldots, k\}$ containing m elements, and the sum in (ii) is over all disjoint subsets $(i_1, \ldots, i_m)$, $(j_1, \ldots, j_m)$ of $\{1, \ldots, k\}$ each containing m elements.

Proof of Lemma 3.2

(i) $$\sum_{(i_1,\ldots,i_m)} \left(\sum_{r=1}^{m} a_{i_r} - m\bar{a} \right)^2$$

$$= \sum_{(i_1,\ldots,i_m)} \left[\left(\sum_{r=1}^{m} a_{i_r}\right)^2 - 2m\bar{a} \sum_{r=1}^{m} a_{i_r} + m^2 \bar{a}^2 \right]$$

$$= \sum_{(i_1,\ldots,i_m)} \left(\sum_{r=1}^{m} a_{i_r}^2 + m^2 \bar{a}^2 - 2m\bar{a} \sum_{r=1}^{m} a_{i_r} + \sum_{\substack{p \neq q\\p,q=1}}^{m} a_{i_p} a_{i_q} \right)$$

$$= \frac{mC_m^k}{k} \sum_{i=1}^{k} a_i^2 + m^2 C_m^k \bar{a}^2 - 2m\bar{a} \frac{mC_m^k}{k} \sum_{i=1}^{k} a_i + \sum_{(i_1,\ldots,i_m)} \sum_{\substack{p \neq q\\p,q=1}}^{m} a_{i_p} a_{i_q}$$

$$= \frac{mC_m^k}{k} \sum_{i=1}^{k} a_i^2 - m^2 C_m^k \bar{a}^2 + \frac{C_m^k (m^2 - m)}{k^2 - k} \sum_{i \neq j} a_i a_j$$

$$= \frac{mC_m^k}{k} \sum_{i=1}^{k} a_i^2 - m^2 C_m^k \bar{a}^2 + \frac{C_m^k (m^2 - m)}{k^2 - k} \left(k^2 \bar{a}^2 - \sum_{i=1}^{k} a_i^2 \right)$$

$$= \frac{m(k-m) C_m^k}{k(k-1)} \left(\sum_{i=1}^{k} a_i^2 - k\bar{a}^2 \right)$$

$$= \frac{m(k-m) C_m^k}{k(k-1)} \sum_{i=1}^{k} (a_i - \bar{a})^2 .$$

(ii) $$\sum_{\substack{(i_1,\ldots,i_m)\\(j_1,\ldots,j_m)}} \left(\sum_{r=1}^{m} a_{i_r} - \sum_{s=1}^{m} a_{j_s} \right)^2$$

$$= \sum_{\substack{(i_1,\ldots,i_m)\\(j_1,\ldots,j_m)}} \left[\left(\sum_{r=1}^{m} a_{i_r}\right)^2 + \left(\sum_{s=1}^{m} a_{i_s}\right)^2 - 2 \sum_{r=1}^{m} \sum_{s=1}^{m} a_{i_r} a_{j_s} \right]$$

$$= \sum_{\substack{(i_1,\ldots,i_m)\\(j_1,\ldots,j_m)}} \left[2\left(\sum_{r=1}^{m} a_{i_r}\right)^2 - 2 \sum_{r=1}^{m} \sum_{s=1}^{m} a_{i_r} a_{j_s} \right]$$

$$= 2C_m^{k-m} \sum_{(i_1,\ldots,i_m)} \left(\sum_{r=1}^{m} a_{i_r}\right)^2 - 2 \sum_{\substack{(i_1,\ldots,i_m)\\(j_1,\ldots,j_m)}} \sum_{r=1}^{m} \sum_{s=1}^{m} a_{i_r} a_{j_s}$$

$$= 2C_m^{k-m} \sum_{(i_1,\ldots,i_m)} [\sum_{r=1}^{m} a_{i_r}^2 + \sum_{r \neq s} a_{i_r} a_{i_s}] - \frac{2C_m^k C_m^{k-m} m^2}{k(k-1)} \sum_{i \neq j} a_i a_j$$

$$= 2C_m^{k-m} [\frac{C_m^k m}{k} \sum_{r=1}^{m} a_{i_r}^2 + \frac{C_m^k m(m-1)}{k(k-1)} \sum_{i \neq j} a_i a_j] - \frac{2C_m^k C_m^{k-m} m^2}{k(k-1)} \sum_{i \neq j} a_i a_j$$

$$= \frac{2mC_m^k C_m^{k-m}}{k-1} [\frac{k-1}{k} \sum_{i=1}^{k} a_i^2 - \frac{1}{k} \sum_{i \neq j} a_i a_j]$$

$$= \frac{2mC_m^k C_m^{k-m}}{k-1} [\frac{k-1}{k} \sum_{i=1}^{k} a_i^2 - \frac{1}{k} (k^2 \bar{a}^2 - \sum_{i=1} a_i^2)]$$

$$= \frac{2mC_m^k C_m^{k-m}}{k-1} (\sum_{i=1}^{k} a_i^2 - k\bar{a}^2)$$

$$= \frac{2mC_m^k C_m^{k-m}}{k-1} \sum_{i=1}^{k} (a_i - \bar{a})^2 \ \#$$

We now state and prove Corollary 3.4

Corollary 3.4 Consider the maximin test given in equation (3.9). If under H_0

$P\{ \sum_J \cosh\,[w(\bar{x}_J - \bar{x})] \geq c_\delta \} = \alpha, \quad \delta \in (0, \infty)$, and

$P\{ \sum_{i=1}^{k} (x_i - \bar{x})^2 \geq c_1 \} = \alpha, \ P\{ \max_J |\bar{x}_J - \bar{x}| \geq c_2 \} = \alpha.$

Then we have

(i) $\lim_{\delta \to 0} I_{(\sum_J \cosh\,[w(\bar{x}_J - \bar{x})] < c_\delta)}(\mathbf{x}) \overset{a.s.}{=} I_{(\sum_{i=1}^{k} (x_i - \bar{x})^2 < c_1)}(\mathbf{x})$,

(ii) $\lim_{\delta \to +\infty} I_{(\sum_J \cosh\,[w(\bar{x}_J - \bar{x})] < c_\delta)}(\mathbf{x}) \overset{a.s.}{=} I_{(\max_J |\bar{x}_J - \bar{x}| < c_2)}(\mathbf{x})$,

where the maximum is taken over all subsets J of $\{1,\ldots,k\}$ each of size j, and $I_A(.)$ is the index function of the set A.

Proof of Corollary 3.4

Notice that

$$V_\delta \equiv \{ \sum_J \cosh\{ w(\bar{x}_J - \bar{x}) \} < c_\delta \}$$

$$= \{ \sum_J [2 + w^2 (\bar{x}_J - \bar{x})^2] + \sum_J [\cosh\{ w(\bar{x}_J - \bar{x}) \} - 2 - w^2 (\bar{x}_J - \bar{x})^2] < c_\delta \}$$

$$= \{ \sum_J (\bar{x}_J - \bar{x})^2 + \frac{1}{w^2} \sum_J [\cosh\{ w(\bar{x}_J - \bar{x})\} - 2 - w^2(\bar{x}_J - \bar{x})^2] < \frac{c_\delta - 2C_j^k}{w^2} \}$$

$$= \{ \sum_{j=1}^{k} (x_j - \bar{x})^2 + \frac{1}{C_0 w^2} \sum_J [\cosh\{ w(\bar{x}_J - \bar{x})\} - 2 - w^2(\bar{x}_J - \bar{x})^2] < \frac{c_\delta - 2C_j^k}{C_0 w^2} \}$$

$$= \{ \sum_{j=1}^{k} (x_j - \bar{x})^2 + \frac{1}{C_0 w^2} \sum_J f(\, w(\bar{x}_J - \bar{x})\,) < d_\delta \} , \qquad (**)$$

where C_0 is a positive constant (which is independent of δ and determined in Lemma 3.2), and

$$f(x) = cosh(x) - 2 - x^2 \ , \quad d_\delta = \frac{c_\delta - 2C_j^k}{C_0 w^2} \ .$$

To prove the first statement of the corollary, we first need to prove that $\lim_{\delta \to 0} d_\delta = c_1$. If this is not true, then there exists a set $\{\delta_n\}$ such that $\delta_n \downarrow 0$ as $n \to \infty$, and $\lim_{n\to\infty} d_{\delta_n} = d \neq c_1$. Without loss of generality, we suppose that $d = c_1 + r > c_1$. Then for any $\mathbf{x}$ satisfying $\sum_{j=1}^{k} (x_j - \bar{x})^2 \le c_1 + r/4$, we have

$$\sum_{j=1}^{k} (x_j - \bar{x})^2 + \frac{1}{C_0 w^2} \sum_J f(\, w(\bar{x}_J - \bar{x})\,)$$

$$\le \sum_{j=1}^{k} (x_j - \bar{x})^2 + \frac{1}{C_0 w^2} \sum_J f(\, w\sqrt{c_1 + r/4}\,)$$

$$\le c_1 + r/4 + \frac{C_j^k}{C_0 w^2} \ f(\, w\sqrt{c_1 + r/4}\,)$$

$$\le c_1 + r/4 + r/4$$

$$\le d_{\delta_n} \ , \quad \text{for sufficient large } n \ ,$$

where the first and third inequalities follow from the fact that $f(x)$ increases as $|x|$ increases and $\lim_{x\to 0} f(x)/x^2 = 0$, and the fourth inequality follows from the assumption that $\lim_{n\to\infty} d_{\delta_n} = c_1 + r$. So, we have in fact proved that, for sufficient large n,

$$\{ \sum_{j=1}^{k} (x_j - \bar{x})^2 \le c_1 + r/4 \} \subset \{ \sum_J \cosh[w(\bar{x}_J - \bar{x})] \le c_{\delta_n} \} ,$$

which leads to a contradiction since, under H_0, the probability of the left set is larger than $1-\alpha$ whilst the right set always has probability equal to $1-\alpha$. Thus we have proved that $\lim_{\delta \to 0} d_\delta = c_1$. Now we can see, from the expression (**), the property of $f(x)$ and the fact that $\lim_{\delta \to 0} d_\delta = c_1$, that

$$\lim_{\delta \to 0} I_{(\sum_J \cosh [w(\bar{x}_J - \bar{x})] < c_\delta)}(\mathbf{x}) = I_{(\sum_{i=1}^{k}(x_i - \bar{x})^2 < c_1)}(\mathbf{x}) \, ,$$

for any $\mathbf{x}$ satisfying that $\sum_{i=1}^{k}(x_i - \bar{x})^2 \neq c_1$. Since

$$P(\sum_{i=1}^{k}(x_i - \bar{x})^2 = c_1) = 0,$$

the first statement has been proved. To prove the second part of the corollary, notice that

$$V_\delta \equiv \{ \sum_J \cosh\{w(\bar{x}_J - \bar{x})\} \le c_\delta \}$$

$$= \{ \max_J |\bar{x}_j - \bar{x}| + \frac{1}{w} ln(B) \le \tilde{d}_\delta \} \, ,$$

where

$$B = \sum_J \{ \exp(-w[\max_J |\bar{x}_J - \bar{x}| + (\bar{x}_J - \bar{x})]) + \exp(-w[\max_J |\bar{x}_J - \bar{x}| - (\bar{x}_J - \bar{x})]) \}$$

and

$$1 \le B \le \sum_J 2 = 2\, C_j^k \, , \quad \text{for any } \mathbf{x}, \text{ and } w > 0 \, .$$

Similar arguments as the above can then be applied to yield the second statement. The detail is omitted here #

It is clear that as δ approaches zero, the power level of the maximin test at the least favourable configuration approaches the size α, and as δ approaches infinity, the power level approaches one. Therefore, for the difference zone given in equation (3.1), it is clear that a test based on the range statistic $\max_{1 \le i \le K} |x_i - \bar{x}|$ is asymptotically optimal as the power level approaches one, while a test based on the sum of squares statistic $\sum_{i=1}^{k}(x_i - \bar{x})^2$ is asymptotically optimal as the power level approaches the size α. A comparison of the least favourable power of these two tests for different values of δ, when $\alpha = 0.05$, $k = 3$ is presented in Table 3.1. Clearly for the case considered, the test based on the sum of squares is more powerful than the test based on the range statistic only when δ is very small (the corresponding powers are less than 0.06), for other values of δ, the powers of the range based test dominate the powers of the sum of squares based test.

Next in Theorem 3.5 we derive the maximin test for difference zones of the form $\Theta_A(c_2(j),\delta)$.

Theorem 3.5 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ with size $\alpha \in (0, 1)$, the maximin test $\psi^*(\mathbf{x})$ for the difference zone $\Theta_A(c_2(j),\delta)$ is given by

$$\psi^*(\mathbf{x}) = 1 \quad \Longleftrightarrow \quad \sum_{J_1,J_2} \cosh\{z(\bar{x}_{J_1} - \bar{x}_{J_2})\} \geq c \tag{3.13}$$

where the sum is over all disjoint subsets J_1, J_2 of $\{1,\ldots,k\}$ each containing j elements, $\bar{x}_{J_i}$ is the average of the element of $\mathbf{x}$ with indices in J_i,

$$z = \frac{\delta}{2\sigma^2(1-\rho)}$$

and the constant c is chosen so that the test has size α.

Proof of Theorem 3.5

Since $\mathbf{1}'\mu_2(j) = 0$, it follows from Theorem 3.2 and arguments similar to those at the beginning of the proof of Theorem 3.4 that for the difference zone $\Theta_A(c_2(j),\delta)$, the maximin test is of the form

$$\psi^*(\mathbf{x}) = 1 \quad \Longleftrightarrow \quad \sum_{\pi\in\Pi} \cosh\{(\pi\mathbf{x})' \Sigma^{-1} \mu_2(j)\} \geq c', \tag{3.14}$$

for a suitable constant c'. Now since

$$\Sigma^{-1}\mu_2(j) = \frac{1}{\sigma^2(1-\rho)} \mu_2(j) ,$$

it is clear that the tests given in equations (3.13) and (3.14) are equivalent, and this completes the proof of Theorem 3.5 #

Again, we can see from Theorem 3.5 that the maximin test depends on δ, but depends on the size α only through the critical point. Also, for the difference zone given in equation (3.2), it can be seen that we have the following corollary.

Corollary 3.5 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ with size $\alpha \in (0, 1)$, the maximin test $\psi^*(\mathbf{x})$ for the difference zone $\{ \mu: \max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge \delta \}$ is given by

$$\psi^*(\mathbf{x}) = 1 \iff \sum_{i=1}^{k} \cosh\{z'(x_i - x_j)\} \ge c \tag{3.15}$$

where z' is equal to z in Theorem 3.4, and the constant c is chosen to give the required size.

In Corollary 3.6 below we consider what happens to the maximin test as δ approaches zero or infinity; as δ approaches zero, we again see that it is equivalent to a test based on the sum of squares statistic, and as δ approaches infinity, the test is equivalent to a test based on a range statistic

Corollary 3.6 Consider the maximin test given in equation (3.13). If under H_0

$P\{ \sum_{J_1, J_2} \cosh\, [z(\bar{x}_{J_1} - \bar{x}_{J_2})] \ge c_\delta \} = \alpha$, $\delta \in (0, \infty)$, and

$P\{ \sum_{i=1}^{k} (x_i - \bar{x})^2 \ge c_1 \} = \alpha$, $P\{ \max_{J_1, J_2} |\bar{x}_{J_1} - \bar{x}_{J_2}| \ge c_2 \} = \alpha$.

Then we have

(i) $$\lim_{\delta \to 0} I_{(\sum_{J_1 J_2} \cosh\, [z(\bar{x}_{J_1} - \bar{x}_{J_2})] < c_\delta)}(\mathbf{x}) \overset{a.s.}{=} I_{(\sum_{i=1}^{k} (x_i - \bar{x})^2 < c_1)}(\mathbf{x}),$$

(ii) $$\lim_{\delta \to +\infty} I_{(\sum_{J_1 J_2} \cosh\, [z(\bar{x}_{J_1} - \bar{x}_{J_2})] < c_\delta)}(\mathbf{x}) \overset{a.s.}{=} I_{(\max_{J_1 J_2} |\bar{x}_{J_1} - \bar{x}_{J_2}| < c_2)}(\mathbf{x}),$$

where the maximum is taken over all disjoint subsets J_1, J_2 of $\{1, \ldots, k\}$ each of size j, and $I_A(.)$ is the index function of the set A.

Proof of Corollary 3.6

As the proof is similar to that of Corollary 3.4, it is omitted #

Finally then, for the difference zone given in equation (3.2), a test based on the range statistic $\max_{1 \le i,j \le k} |x_i - x_j|$ will be asymptotically optimal as the power level approaches one, and a test based on the sum of squares statistic $\sum_{i=1}^{k} (x_i - \bar{x})^2$ will be asymptotically optimal as the power level approaches the size α. We can again make a comparison of the least favourable powers of these two asymptotically optimal tests for different values of δ, when $\alpha = 0.05$, and $k = 8$. The results are presented in Table 3.2. Clearly when δ is small (the corresponding powers are less than 0.33), the test based on the sum of squares is more powerful than the test based on the range statistic, while for the larger values of δ, the range based test is always more powerful. The difference between the least favourable powers of the range based test and the sum of squares based test can be as much as 2.29%.

It may well be the case that in defining a difference zone, the experimenter is quite clear as to which variability measure function $b(\mu)$ will provide a difference zone with a useful shape, but is not so clear in the choice of δ. The choice of δ is important since the optimal tests given in equations (3.9) and (3.13) depend on its value. One way to avoid this problem is for the experimenter to specify a non-negative function $h(\delta)$, $\delta \geq 0$, and then to use a test procedure which maximizes the worst power within the difference zone averaged over all values of δ, $0 \leq \delta < \infty$, weighted with respect to the function $h(\delta)$. Thus, if we suppose that $\mu(\delta)$ is a least favourable configuration for the difference zone $\Theta_A(\mathbf{c},\delta)$ for all tests $\psi(\mathbf{x}) \in \Psi$, then the optimal test will be the test $\psi(\mathbf{x}) \in \Psi$ which maximizes

$$\int_{\delta=0}^{\infty} \int_{\mathbf{R}^k} \psi(\mathbf{x})\, \varphi(\mathbf{x}-\mu(\delta))\, h(\delta)\, d\mathbf{x} d\delta \tag{3.16}$$

subject to the size constraint.

Theorem 3.6 below gives the optimal test for the difference zones of the form $\Theta_A(\mathbf{c}_1(j),\delta)$.

Theorem 3.6 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ with size $\alpha \in (0, 1)$, the optimal test $\psi^*(\mathbf{x})$ for the difference zone $\Theta_A(\mathbf{c}_1(j),\delta)$, which maximizes the expression in (3.16), is given by

$$\psi^*(\mathbf{x}) = 1 \quad \Longleftrightarrow$$

$$\sum_{\pi \in \Pi} \int_{\delta=0}^{\infty} exp(-\frac{kb}{2j(k-j)}\delta^2) \cosh \{\, b(\pi\mathbf{x})'\mu_1(j) \,\}\, h(\delta)\, d\delta \geq c \tag{3.17}$$

where $b = \dfrac{1}{\sigma^2(1-\rho)}$, $\mu_1(j)$ is defined in (3.4) and the constant c is chosen so that the test has size α.

Proof of Theorem3.6

The optimal test $\psi^*(\mathbf{x})$ is the test which maximizes

$$L(\Psi) = \int_{\delta=0}^{\infty} \int_{\mathbf{R}^k} \psi(\mathbf{x})\ \varphi(\mathbf{x}-\mu_1(j))\ h(\delta)\ d\mathbf{x}d\delta$$

over all tests $\psi(\mathbf{x}) \in \Psi$ subject to the size condition that

$$\alpha = \int_{\mathbf{R}^k} \psi(\mathbf{x})\ \varphi(\mathbf{x})\ d\mathbf{x}\ .$$

It follows arguments similar to the proof of Theorem 3.3 that the optimal test is of the form

$$\psi^*(\mathbf{x}) = 1 \quad \Longleftrightarrow \quad \frac{G(\mathbf{x})}{\varphi(\mathbf{x})} \geq c$$

where, in this case,

$$G(\mathbf{x}) = \sum_{\pi\in\Pi} \int_{\delta=0}^{\infty} [\ \varphi(\pi\mathbf{x}-\mu_1(j)) + \varphi(\pi\mathbf{x}+\mu_1(j)\ ]\ h(\delta)\ d\delta,$$

provided that this test is within the class Ψ. However,

$$\frac{G(\mathbf{x})}{\varphi(\mathbf{x})} \propto \sum_{\pi\in\Pi} \int_{\delta=0}^{\infty} \exp(-\frac{b}{2}\mu_1(j)'\mu_1(j))\ \cosh\{\ b(\pi\mathbf{x})'\mu_1(j)\ \}\ h(\delta)\ d\delta$$

where $b = \dfrac{1}{\sigma^2(1-\rho)}$, and so it is clear that this test satisfies the requirements to be in the class Ψ. Thus the proof is completed #

For a general function $h(\delta)$, expression (3.17) may be used to find the optimal test for difference zones of the form $\Theta_A(\mathbf{c}_1(j),\delta)$. In particular, if the weight is only put at one point $\delta \in (0, \infty)$, then the optimal test becomes the maximin test for the difference zone $\Theta_A(\mathbf{c}_1(j),\delta)$. If the weight function $h(\delta) = \exp\{-\delta^2/2\tau^2\}$, then the optimal test reduces to

$$\psi^*(\mathbf{x}) = 1 \quad \Longleftrightarrow \quad \sum_{\pi\in\Pi} \exp\left\{\frac{b^2((\pi\mathbf{x})'\mu)^2}{2(b\mu'\mu+\frac{1}{\tau^2})}\right\} \geq c \qquad (3.18)$$

for suitable choice of the constant c, and

$$\mu = (\ -\frac{1}{j},\ \dots\ ,\ -\frac{1}{j},\ \frac{1}{k-j},\ \dots\ ,\ \frac{1}{k-j}\ )$$

where $-1/j$ occupies the first j places. For a constant weight function $h(\delta)$ we can obtain the optimal test by taking τ to be infinite in equation (3.18). Thus, if we wish to maximize the average power uniformly over all values of δ

for the difference zone given in equation (3.1), we have the following Corollary concerning the form of the optimal test.

Corollary 3.7 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$, and $h(\delta) = constant$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ with size $\alpha \in (0, 1)$, the optimal test $\psi^*(\mathbf{x})$ for the difference zone $\{ \mu : \max_{1 \le i \le k} |\mu_i - \bar{\mu}| \ge \delta \}$ is given by

$$\psi^*(\mathbf{x}) = 1 \iff \sum_{i=1}^{k} \exp\left\{ \frac{bk}{2(k-1)} (x_i - \bar{x})^2 \right\} \ge c \tag{3.19}$$

where $b = \dfrac{1}{\sigma^2(1-\rho)}$ and the constant c is chosen so that the test has size α.

For the difference zones of the form $\Theta_A(c_2(j),\delta)$, we have the theorem analogous to Theorem 3.6.

Theorem 3.7 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ with size $\alpha \in (0, 1)$, the optimal test $\psi^*(\mathbf{x})$ for the difference zone $\Theta_A(c_2(j),\delta)$, which maximizes the expression in (3.16), is given by

$$\psi^*(\mathbf{x}) = 1 \iff$$

$$\sum_{\pi \in \Pi} \int_{\delta=0}^{\infty} exp(-\frac{b}{4j}\delta^2) \cosh \{ b(\pi \mathbf{x})' \mu_2(j) \} \, h(\delta) \, d\delta \ge c \tag{3.20}$$

where $b = \dfrac{1}{\sigma^2(1-\rho)}$, $\mu_2(j)$ is defined in (3.5) and the constant c is chosen so that the test has size α.

Proof of Theorem 3.7

As the proof is similar to that of Theorem 3.6, we do not supply the details here #

Again, expression (3.20) may be used to find, for a general function $h(\delta)$, the optimal test for difference zones of the form $\Theta_A(c_2(j),\delta)$. If the weight is put at only one point $\delta \in (0, \infty)$, then the optimal test again becomes the maximin test for the difference zone $\Theta_A(c_2(j),\delta)$. If the weight function $h(\delta) = exp\ \{-\delta^2/2\tau^2\}$, then the optimal test reduces to

$$\psi^*(\mathbf{x}) = 1 \iff \sum_{\pi \in \Pi} \exp\left\{ \frac{b^2((\pi \mathbf{x})'\mu)^2}{2(b\mu'\mu + \frac{1}{\tau^2})} \right\} \ge c \tag{3.21}$$

for suitable choice of the constant c, and

$$\mu = (-\frac{1}{2j}, \dots , -\frac{1}{2j}, 0, \dots , 0, \frac{1}{2j}, \dots , \frac{1}{2j})$$

where $-\frac{1}{2j}$ and $\frac{1}{2j}$ each occupy j places. For a constant weight function $h(\delta)$ we can obtain the optimal test by taking τ to be infinite in equation (3.21). Thus, if we wish to maximize the average power uniformly over all values of δ for the difference zone given in equation (3.2), the optimal test is given in the following corollary.

Corollary 3.8 Let $f(\mathbf{x}) = \varphi(\mathbf{x})$, and $h(\delta) = constant$. Then within the class of tests $\psi(\mathbf{x}) \in \Psi$ with size $\alpha \in (0, 1)$, the optimal test $\psi^*(\mathbf{x})$ for the difference zone $\{ \mu : \max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge \delta \}$ is given by

$$\psi^*(\mathbf{x}) = 1 \iff \sum_{1 \le i,j \le k} \exp \left\{ \frac{b}{4} (x_i - x_j)^2 \right\} \ge c \tag{3.22}$$

where $b = \dfrac{1}{\sigma^2(1-\rho)}$ and the constant c is chosen so that the test has size α.

From the results of this section we can see that, for the one-way model considered at the beginning of section 3.1, a test based on a range statistic corresponding to the particular difference zone considered is asymptotically optimal as the power level approaches one, while a test based on the sum of squares statistic is asymptotically optimal as the power level approaches the size α for all alternative sets considered.

The one-way model will also usually involve a nuisance parameter σ^2 which may be estimated unbiasedly by the usual analysis of variance mean square statistic S^2 with a chi-square distribution. All of the tests proposed in this chapter may be modified for this case by replacing X_i by X_i/S, and by replacing δ by $\delta\sigma$. Thus a test based on the sum of squares statistic is equivalent to the common F-test, a test based on the range statistic $\max_{1 \le i,j \le k} |x_i - x_j|$ is equivalent to the Studentized range test, and a test based on the range $\max_{1 \le i \le k} |x_i - \bar{x}|$ is equivalent to the comparisons with average test. Thus, in order to guarantee a power requirement for the difference zone given in (3.1) say, our results suggest that the F-test will allow a smaller sample size than the Studentized range test if the required power is small (close to α), but the Studentized range test will permit the use of a smaller sample size than the F-test if the power is large. Such trends can be seen from Table 2.6 in chapter 2. Usually an experiment will be designed to achieve a large power and so it will be better to design an experiment using a range test such as the Studentized range test rather than using the F-test for the difference zones considered. Notice that when the variance σ^2 is unknown, the alternative set Θ_A must be specified in terms of multiples of σ in order to meet the power requirement in a single-stage experiment. If the difference zone is to be specified independently of the unknown variance σ^2, then some kind of two-stage procedure, as proposed in chapter 2, will be necessary in order to meet the power requirement (where examination of the data from the first stage provides information about the variance σ and indicates what sample sizes are required for the second stage).

Test procedures based on the statistics $\max_{1 \le i,j \le k} |x_i - x_j|$, $\max_{1 \le i \le k} |x_i - \bar{x}|$ and $\sum_{i=1}^{k} (x_i - \bar{x})^2$ have an advantage in that when the data is normally distributed, their critical points are well tabulated (see, for example, Hawkins & Perold (1977) and Halperin et al. (1955) for critical points of $\max_{1 \le i \le k} |x_i - \bar{x}|$ based test). Critical points for the optimal tests such as those given in equations (3.12) and (3.15) can be calculated on a computer, in some cases by direct evaluation of the power function, or more generally by simulation. Some calculation results about the critical points and powers of these tests can be seen in Hurn (1989). With the results of this chapter, the willing experimenter will be able to select a test procedure and a sample size to guarantee a stated size and power requirements in an optimal way.

Table 3.1

A comparison of the least favourable powers of the range test $\max_{1 \le i \le k} |x_i - \bar{x}|$ and the sum of squares test $\sum_{i=1}^{k} (x_i - \bar{x})^2$ for the difference zone

$$\Theta_A(c_1(1),\delta) = \{ \mu : \max_{1 \le i \le k} |\mu_i - \bar{\mu}| \ge \delta \},$$

with $k=3$ populations, and size $\alpha=0.05$.

δ	range test	sum of squares test
0.00	0.0500	0.0500
0.05	0.0528	0.0529
0.10	0.0512	0.0511
0.20	0.0551	0.0545
0.30	0.0616	0.0603
0.60	0.0981	0.0929
0.70	0.1163	0.1095
0.80	0.1376	0.1292
0.90	0.1622	0.1521
1.00	0.1901	0.1784
1.30	0.2926	0.2769
1.50	0.3743	0.3572
1.70	0.4630	0.4458
1.90	0.5542	0.5381
2.10	0.6341	0.6291
2.30	0.7253	0.7138
2.60	0.8285	0.8211
2.90	0.9032	0.8991
3.20	0.9509	0.9489
3.50	0.9777	0.9769
3.80	0.9910	0.9907

Table 3.2

A comparison of the least favourable powers of the range test $\max_{1 \le i,j \le k} |x_i - x_j|$

and the sum of squares test $\sum_{i=1}^{k} (x_i - \bar{x})^2$ for the difference zone

$$\Theta_A(c_2(1),\delta) = \{ \mu : \max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge \delta \},$$

with $k=8$ populations, and size $\alpha=0.05$.

δ	range test	sum of squares test
0.00	0.0500	0.0500
0.25	0.0510	0.0511
0.60	0.0559	0.0565
1.10	0.0712	0.0729
1.70	0.1070	0.1102
2.40	0.1830	0.1857
2.85	0.2557	0.2559
3.20	0.3254	0.3221
3.60	0.4168	0.4082
4.00	0.5160	0.5017
4.45	0.6288	0.6090
4.90	0.7333	0.7104
5.40	0.8302	0.8075
5.85	0.8956	0.8760
6.25	0.9367	0.9211
6.55	0.9584	0.9459
6.80	0.9715	0.9616
7.55	0.9922	0.9881
8.05	0.9971	0.9952
8.40	0.9987	0.9976
9.05	0.9997	0.9994

Chapter 4

Testing the equality of several proportions

4.1 Introduction

In this chapter the elementary problem of testing the equality of several Bernoulli probabilities is considered, and attention is especially directed to forming an assessment of power properties of such a test. In particular, power levels are considered under a simple restriction on the range of the Bernoulli probabilities. A test procedure based on the range of the arcsin-root transformations of the observed proportions is considered, and it is shown how power levels may be calculated both exactly and under asymptotic assumptions.

Suppose that we have k independent treatments with Bernoulli responses in terms of their success probabilities $\pi_1, \ldots, \pi_k$, and that our data consist of n independent observations on each of the k treatments. Let Y_i denotes the number of successes from the ith treatment, the sample proportion of successes being $\hat{\pi}_i = Y_i/n$ $(1 \leq i \leq k)$. Then, the elementary statistical problem is to test the null hypothesis $H_0 : \pi_1 = \ldots = \pi_k$ of equality of the k success probabilities against a general two-sided alternative hypothesis "not H_0". Many simultaneous test procedures have been proposed for this problem in the literature, most of which are based on large sample normal approximation, and concentrate on guaranteeing the probability of type I error asymptotically as the sample size n becomes large, since even for making a single inference very few exact small sample procedures are available. One usual test procedure for this problem, assuming a large sample approximation to $\hat{\pi}_i$

$$\hat{\pi}_i \sim N(\, \pi_i, \frac{\pi_i(1-\pi_i)}{n}), \qquad 1 \le i \le k,$$

is based on the Tukey-Kramer (TK) procedure (see Hochberg & Tamhane(1987)), which employs the test statistic

$$Q' = \max_{1 \le i,j \le k} \frac{|\, \hat{\pi}_i - \hat{\pi}_j \,|}{\sqrt{\frac{1}{2}\left[\frac{\hat{\pi}_i(1-\hat{\pi}_i)+\hat{\pi}_j(1-\hat{\pi}_j)}{n}\right]}}$$

and rejects the null hypothesis H_0 iff $Q' \ge q^{\alpha}_{k,\infty}$, where $q^{\alpha}_{k,\infty}$ is the upper α-point of the Studentised range distribution with parameters k and ∞. This test will be asymptotically of size α. However, it is not easy to evaluate the exact size and the exact power of this test procedure when $K \ge 4$. Goodman (1964) derived simultaneous confidence intervals for contrasts among several multinomial populations using large sample χ^2 approximation. Bhaphar & Somes (1976) considered the problem for matched samples. Knoke (1976) considered a so-called maximal contrast test that is essentially the Scheffe procedure applied to dichotomous data; using Monte Carlo simulations he studied the significance levels and powers of several competing procedures to test the null hypothesis H_0. More recently, Williams (1988) investigated the small sample behaviour of various such test procedures; he pointed out that it is now feasible to evaluate the exact significance probability of tests on Bernoulli data and it is no longer necessary to rely on asymptotic approximation. We, here, consider in addition the problem of forming an assessment of the power of the test procedure employed.

For the classical problem of testing the equality of k normal means in a one-way layout, it has been suggested that the power level be guaranteed when the range of the means exceeds a certain value, and it seems generally agreed that such an approach affords the experimenter an intuitive interpretation of the sensitivity of an experiment. Therefore, for the problem at hand, we consider the following approach: for a test of the null hypothesis H_0 of specified size α, guarantee the power requirement that the probability of rejecting H_0 is no less than β whenever

$$\max_{1 \le i,j \le k} |\pi_i - \pi_j| \ge \delta \qquad (4.1)$$

for specified values of $\beta > \alpha$ and $\delta > 0$. The use of this approach allows the experimenter to gauge the sensitivity of the experiment in terms of the differences between the unknown success probabilities.

Using a simple test based on the range of the $\arcsin\sqrt{\hat{\pi}_i}$, $1 \le i \le k$, we show how these probability requirements may be satisfied under the assumption of the asymptotic ($n \to \infty$) normal distribution

$$\arcsin\sqrt{\hat{\pi}_i} \sim N(\arcsin\sqrt{\pi_i},\ 1/4n) \qquad 1 \le i \le k. \tag{4.2}$$

This common transformation (see, for example, Bishop et al. (1975) section 14.6.2) is used because it uniquely possesses the property of stabilizing the asymptotic variances to be independent of the probabilities π_i, which considerably simplifies the problem of assessing the power properties of the test procedure. For given values of α, β and δ, we show how the sample size n may be chosen to guarantee the probability requirements. Conversely, for a given sample size n, it is possible to calculate the power level β achieved for different values of α and δ. We also consider the exact size and exact power probabilities of our procedure. We show how the exact size and power may be calculated for our test procedure and provide some illustrative tables of the power levels afforded by certain sample sizes.

We propose our test procedure based on the range of $\arcsin\sqrt{\hat{\pi}_i}$, rather than say a test based on the sum of squared differences of $\arcsin\sqrt{\hat{\pi}_i}$ which would asymptotically have a chi-squared distribution, because a range test is generally good at guaranteeing power levels under the range restrictions such as (4.1) (see chapter 3), and because the exact size and power levels are easily obtained for this test procedure.

4.2 Test procedure and power assessment

If the experimenter is interested in determining whether the success probabilities of the k treatments are identical, then it is appropriate to test the null hypothesis H_0. Let the difference zone $\Theta_A(\delta)$ be defined by

$$\Theta_A(\delta) = \{\ \boldsymbol{\pi}=(\pi_1,\ \dots\ ,\ \pi_k) : \max_{1 \le i,j \le k} |\pi_i - \pi_j| \ge \delta\ \} \subset [0,1]^k$$

for some positive constant δ. The problem is then to find a size α test of the null hypothesis H_0 which satisfies the following power requirement

$$\pi \in \Theta_A(\delta) \quad \Rightarrow \quad P(\text{ reject } H_0\ ) \geq \beta \tag{4.3}$$

for a fixed power level β.

We propose the following test procedure. Using the notation of section 4.1 define

$$Q \;=\; 2\sqrt{n}\ (\ \max_{1 \leq i,j \leq k} |\ \arcsin\sqrt{\hat{\pi}_i} - \arcsin\sqrt{\hat{\pi}_j}\ |\ ),$$

and reject the null hypothesis H_0 iff $Q > q^{\alpha}_{k,\infty}$, where $q^{\alpha}_{k,\infty}$ is the upper α-point of the Studentised range distribution with parameters k and ∞. As a consequence of the asymptotic properties of the arcsin-root transformation given in equation (4.2), the asymptotic distribution of Q under the null hypothesis H_0 is independent of the common value taken by the probabilities π_i, and the test procedure is asymptotically of size α. Notice, however, that for a fixed sample size n, the distribution of Q under the null hypothesis H_0 depends on the particular value taken by the probabilities π_i. Also, in practice, it seems that for small sample sizes, a variance stabilizing correction of Anscombe (1948) is helpful in controlling the exact size of the test procedure, so that it is better to use $\hat{\pi}_i = (Y_i+3/8)/(n+3/4)$ in the test statistic rather than the actual exact sample proportion, and also $\sqrt{n+1/2}$ in place of $\sqrt{n}$.

We now address the problem of determining the smallest sample size n which will achieve the power requirement (4.3) assuming the asymptotic distribution of Q. Let

$$g(\pi,n) = P\{\ Q > q^{\alpha}_{k,\infty}\ |\ \pi,n\ \}$$

be the probability of rejecting the null hypothesis for given values of $\pi = (\pi_1, \ldots, \pi_k)$ and n. Then the asymptotic least favourable configurations of success probabilities π are defined to be those values of $\pi \in \Theta_A(\delta)$ which minimize the power function $g(\pi,n)$ assuming the asymptotic distribution for Q. The following lemma identifies an asymptotic least favourable configuration.

Lemma 4.1 Let $\delta \in (0,1)$, and define $\pi^*(\delta) = (\ (1-\delta)/2, 1/2, \ldots, 1/2, (1+\delta)/2\ )$. Then assuming the asymptotic distribution of Q, we have for any sample size n that

$$\pi \in \Theta_A(\delta) \quad \Rightarrow \quad g(\pi,n) \geq g(\pi^*(\delta),n).$$

Proof of Lemma 4.1

Without loss of generality assume that $\pi_1 \le \ldots \le \pi_k$. Notice that

$$\pi_k - \pi_1 = a \quad \Rightarrow \quad \arcsin\sqrt{\pi_k} - \arcsin\sqrt{\pi_1} \ge b$$

where

$$b = \arcsin\sqrt{\frac{1+a}{2}} - \arcsin\sqrt{\frac{1-a}{2}},$$

an increasing function in a, with equality achieved only when $\pi_1 = (1-a)/2$ and $\pi_k = (1+a)/2$. In addition, it follows from Corollary 3.2 of chapter 3 that, assuming asymptotic normal distributions (4.2), under the condition

$$\max_{1 \le i,j \le k} |\arcsin\sqrt{\pi_i} - \arcsin\sqrt{\pi_j}| \ge b,$$

the power function $g(\pi,n)$ is minimised when

$$\begin{aligned} \arcsin\sqrt{\pi_1} &= c - b/2 \\ \arcsin\sqrt{\pi_i} &= c \qquad\qquad 2 \le i \le k-1 \\ \arcsin\sqrt{\pi_k} &= c + b/2, \end{aligned}$$

for any constant c, and the value of the power function in this case is increasing with b. Consequently, under the condition (4.1), the power function $g(\pi,n)$ is minimized when $\pi_1 = (1-\delta)/2$, $\pi_k = (1+\delta)/2$ and $\pi_i = 1/2$ for $2 \le i \le k-1$. This completes the proof of the lemma #

As a consequence of the lemma, in order to satisfy the power requirement (4.3) assuming the asymptotic distribution of Q, it is necessary and sufficient to choose the sample size n so that $g(\pi^*(\delta),n) \ge \beta$. The evaluation of $g(\pi^*(\delta),n)$ assuming the asymptotic distribution of Q is simple and requires only the evaluation of a one-dimensional integral regardless of k, the number of populations (see equation (2.19) of chapter 2), and a search may easily be made by computer to find the required sample size n.

However, since a computer program is needed anyway to evaluate the power function $g(\pi^*(\delta),n)$ even assuming the asymptotic distribution of Q, it may be just as convenient to evaluate the power function exactly. The exact value of $g(\pi,n)$, assuming that the variance stabilising correction is applied, can be evaluated using the following expression with $c = q^{\alpha}_{k,\infty}/2\sqrt{n+1/2}$.

$$1 - g(\pi,n) =$$

$$P\{ \max_{1\le i\le k} \arcsin\sqrt{(Y_i+3/8)/(n+3/4)} - \min_{1\le i\le k} \arcsin\sqrt{(Y_i+3/8)/(n+3/4)} \le c \}$$

$$= \sum_{x=0}^{n} P(\min_{1\le i\le k} Y_i = x,\ \max_{1\le i\le k} Y_i \le U(x))$$

$$= \sum_{x=0}^{n} [P(x \le \min_{1\le i\le k} Y_i,\ \max_{1\le i\le k} Y_i \le U(x))$$

$$- P(x < \min_{1\le i\le k} Y_i,\ \max_{1\le i\le k} Y_i \le U(x))]$$

$$= \sum_{x=0}^{n} [\prod_{i=1}^{k} P(x \le Y_i \le U(x)) - \prod_{i=1}^{k} P(x < Y_i \le U(x))] , \qquad (4.4)$$

where

$$U(x) = (n+3/4) \left[\sin(\min(\arcsin\sqrt{(x+3/8)/(n+3/4)}+c,\pi/2)) \right]^2 - 3/8,$$

and $Y_i \sim B(n,\pi_i)$, $1 \le i \le k$. The difficulty of the exact evaluation of the power function $g(\pi^*(\delta),n)$ is independent of k (since $\pi_2 = \dots = \pi_{k-1}$ and so $k-2$ of the product terms in equation (4.4) are identical) but is increasing in the sample size n. In fact

$$g(\pi^*(\delta),n) =$$

$$1 - \sum_{x=0}^{n} [P(x\le Y_1\le U(x))\ P(x\le Y_2\le U(x))^{k-2}\ P(x\le Y_k\le U(x))$$

$$- P(x<Y_1\le U(x))\ P(x<Y_2\le U(x))^{k-2}\ P(x<Y_k\le U(x))] ,$$

where $Y_1 \sim B(n, (1-\delta)/2)$, $Y_2 \sim B(n, 1/2)$ and $Y_k \sim B(n, (1+\delta)/2)$. The exact size of the test procedure when $\pi_i = \bar{\pi}$, $1 \le i \le k$, may also be evaluated using expression (4.4), which becomes

$$g(\pi,n) =$$

$$1 - \sum_{x=0}^{n} \{ [P(x\le Y_1\le U(x))]^k - [P(x<Y_1\le U(x))]^k \} ,$$

where $Y_1 \sim B(n, \bar{\pi})$.

In Table 4.1 we present some calculations of the exact size of the test procedure for a nominal size value of $\alpha = 0.10$, for $k = 2, \dots , 6$ populations, $\pi_i = \bar{\pi} = 0.1, 0.2, \dots , 0.5$, $1 \le i \le k$, and various sample sizes n. These values were calculated using the variance stabilizing correction suggested above.

If this correction is not used then the test tends to be predominantly liberal for the smaller values of $\bar{\pi}$. It can be seen from Table 4.1 that the convergence of the exact size to the nominal size is not monotonic in the sample size, which is due to the discrete nature of the problem. Of course, because of the discreteness, there will not in general be a critical point which will produce a test of size exactly equal to the given nominal value. The only alternatives to a procedure of this kind which has varying size are a randomized test or a test which is always conservative and may have low power, both of which have their disadvantages.

In Table 4.2 we present some calculations of the exact power evaluated at the asymptotic least favourable configuration $g(\pi^*(\delta),n)$ for a nominal size value of $\alpha = 0.10$, for $k = 2, \ldots, 6$ populations, $\delta = 0.1, 0.2$, and various sample sizes n. For a fixed critical point, the power does not always increase as the sample size increases, but then it must be remembered that the exact size of the test is also dependent on the sample size. Thus, certain care needs to be taken in determining the smallest sample size needed to guarantee a certain power level, and the experimenter should take into consideration both the exact size and exact power level associated with a given sample size. It can be seen from Table 4.2 that in order to obtain reasonable power levels, $\beta \geq 0.95$ for $\delta = 0.2$ say, then quite large sample sizes are required, e.g. $n > 130$ for $k = 2$ and $n > 190$ for $k = 6$. For sample sizes of this order it would appear from Table 4.1 that the exact size is generally well approximated by the nominal size.

If the assumption of the asymptotic distribution of Q is not made, then a practical method of sample size determination would perhaps be to first find a sample size for which the exact power function evaluated at the asymptotic least favourable configuration of probabilities, $g(\pi^*(\delta),n)$, with a critical value of $q^{\alpha}_{k,\infty}$ exceeds the required power level β. Then, for that sample size, the exact size of the test procedure could be evaluated at appropriate values of $\pi = (\bar{\pi}, \ldots, \bar{\pi})$. If the experimenter is not happy with the exact size levels then an alteration can be made in the critical point and the process repeated. Of course, for a given sample size n, there may be a set of probabilities $\pi \in \Theta_A(\delta)$ for which the power is smaller than at $\pi^*(\delta)$, although asymptotically this cannot be the case. The experimenter may feel that it is necessary to check the value of the exact power function $g(\pi,n)$ in the neighborhood of $\pi^*(\delta)$ to ensure that a suitable power level is attained.

However, the power levels attained for probabilities $\pi \in \Theta_A(\delta)$ will generally be substantially higher than those attained at $\pi^*(\delta)$, even for probabilities π for which there is equality in the condition (4.1). For example, when $k = 3$, Table 4.3 compares the exact power level calculated at $\pi^*(\delta) = (\,(1-\delta)/2,\ 1/2,\ (1+\delta)/2\,)$ with the exact power level calculated at $\pi'(\delta) = (\,(1-\delta)/2,\ (1-\delta)/2,\ (1+\delta)/2\,)$, for $\alpha = 0.10$ and $\delta = 0.1, 0.2$. For the cases considered, the power level can be as much as 0.1 higher at $\pi'(\delta)$ than at $\pi^*(\delta)$.

Table 4.4 compares the exact power level at $\pi^*(\delta) = ((1-\delta)/2,\ 1/2,\ (1+\delta)/2)$ with the power level calculated also at $\pi^*(\delta)$ but by using the normal approximation (4.2), for $\alpha = 0.10$ and $\delta = 0.1,\ 0.2$. For the cases considered, it seems that the normal approximations are quite reasonable. In particular, for large powers the approximations fit the exact values quite well.

It may well be the case that, for certain experiments the experimenter know that the probabilities π_i are all large, or are all small. In other words, the experimenter may be able to state a bound L, $1/2 \le L \le 1$, and be confident that $\pi_i \ge L$, $1 \le i \le k$, (or equivalently, $\pi_i \le 1-L$, $1 \le i \le k$). In such a situation, the asymptotic least favourable configuration $\pi^*(\delta)$ is not realistic, and guaranteeing power levels at $\pi^*(\delta)$ will be unnecessarily conservative. A natural alternation for this situation is to define the difference zone $\Theta_A(\delta,L)$ by

$$\Theta_A(\delta,L) = \{\ \pi=(\pi_1,\ \ldots\ ,\ \pi_k)\ :\ \max_{1\le i,j\le k} |\pi_i-\pi_j| \ge \delta,\ \min_{1\le i\le k} \pi_i \ge L\ \}$$

for some positive constant δ, and the power requirement is then

$$\pi \in \Theta_A(\delta,L) \quad \Rightarrow \quad P(\text{ reject } H_0\ ) \ge \beta \tag{4.5}$$

for a fixed power level β. The following lemma identifies an asymptotic least favourable configuration for the difference zone $\Theta_A(\delta,L)$.

Lemma 4.2 Let $\delta \in (0,\ 1)$, $L \in [1/2,\ 1)$, and $\delta + L < 1$. Define $\pi^*(\delta,L) = (\ L,\ a,\ \ldots\ ,\ a,\ (L+\delta)\ )$, where

$$a = [\ sin(\ (\ \arcsin\sqrt{L} + \arcsin\sqrt{L+\delta}\ )/2\ )\ ]^2 .$$

Then assuming the asymptotic distribution of Q, we have for any sample size n that

$$\pi \in \Theta_A(\delta,L) \quad \Rightarrow \quad g(\pi,n) \ge g(\pi^*(\delta,L),\, n).$$

Proof of Lemma 4.2

Without loss of generality assume that $\pi_1 \le \ldots \le \pi_k$. Notice that

$$\pi_k - \pi_1 = a, \text{ and } \pi_1 \ge L \quad \Rightarrow \quad \arcsin\sqrt{\pi_k} - \arcsin\sqrt{\pi_1} \ge b'$$

where

$$b' = \arcsin\sqrt{\frac{1+a}{2}} - \arcsin\sqrt{\frac{1-a}{2}},$$

an increasing function in a, with equality achieved only when $\pi_1 = L$ and $\pi_k = L+\delta$. Then following the same argument as in the proof of Lemma 4.1, we have that, for $\pi \in \Theta_A(\delta,L)$, the power function $g(\pi,n)$ is minimised when $\pi_1 = L$, $\pi_k = L+\delta$ and $\pi_i = a$ for $2 \le i \le k-1$. This completes the proof of the lemma #

As a consequence of this lemma, in order to satisfy the power requirement (4.5) assuming the asymptotic distribution of Q, it is necessary and sufficient to choose the sample size n so that $g(\pi^*(\delta,L),\, n) \ge \beta$. The evaluation of the exact value $g(\pi^*(\delta,L),\, n)$ can also be carried out using equation (4.4) with $Y_1 \sim B(n,\, L)$, $Y_2 \sim B(n,\, a)$ and $Y_k \sim B(n,\, L+a)$.

Above, we have considered how to assess the power levels of a test of the homogeneity of k Bernoulli probabilities under the simple range restriction (4.1). Following classical ideas, we suggest that this provides the experimenter with a convenient first-step interpretation of the sensitivity of the test procedure employed. Nevertheless, the limitations of this approach to power assessment, in conjunction with the limitations of hypothesis testing, must be understood. For example, if $\pi_i - \pi_j \ge \delta$, then it is not guaranteed that populations i and j will be responsible for the rejection of the hypothesis test and that an inference may be drawn that $\pi_i > \pi_j$. In order to guarantee specific inferences of this nature, then some kind of multiple comparison or multiple decision procedure is required (see, for example, Gupta & Panchapakesan (1979)). However, it is common practice to perform a hypothesis test of the homogeneity of Bernoulli probabilities, and it is important for the experimenter to have some indication of the power levels achieved.

We have considered a test procedure based on the range of the arcsin-root transformations of the observed population proportions. Some advantages of this test procedure are that, under asymptotic assumptions, the least favourable configuration of probabilities can be easily identified, and also a convenient expression can be found to evaluate the exact power function. Under asymptotic assumptions, power levels can be found by evaluating just a one-dimensional integral expression. However, in practice, it may be just as easy to calculate the exact power levels afforded by various sample sizes and to calculate the exact size of the test procedure.

Table 4.1

The exact size of the test of the null hypothesis H_0, with k populations, sample size n, $\pi=(\bar{\pi},\ldots,\bar{\pi})$ and a nominal size of $\alpha=0.10$.

n	$\bar{\pi}$	k=2	k=3	k=4	k=5	k=6
70	0.1	0.104	0.089	0.089	0.114	0.095
75	0.1	0.102	0.091	0.093	0.113	0.097
80	0.1	0.102	0.094	0.098	0.111	0.099
85	0.1	0.102	0.097	0.102	0.105	0.101
90	0.1	0.102	0.099	0.105	0.106	0.104
100	0.1	0.012	0.101	0.099	0.109	0.095
110	0.1	0.103	0.103	0.099	0.095	0.098
130	0.1	0.104	0.105	0.099	0.098	0.108
150	0.1	0.100	0.101	0.103	0.102	0.098
170	0.1	0.101	0.100	0.101	0.104	0.099
190	0.1	0.098	0.103	0.103	0.103	0.100
70	0.2	0.099	0.105	0.103	0.099	0.101
75	0.2	0.099	0.096	0.098	0.105	0.100
80	0.2	0.102	0.101	0.098	0.097	0.102
85	0.2	0.101	0.099	0.102	0.101	0.105
90	0.2	0.100	0.101	0.104	0.105	0.098
100	0.2	0.099	0.097	0.098	0.100	0.105
110	0.2	0.101	0.101	0.099	0.099	0.099
130	0.2	0.102	0.101	0.099	0.099	0.102
150	0.2	0.100	0.099	0.099	0.100	0.099
170	0.2	0.099	0.099	0.101	0.100	0.101
190	0.2	0.099	0.102	0.099	0.099	0.101
70	0.3	0.097	0.098	0.099	0.097	0.104
75	0.3	0.097	0.101	0.098	0.099	0.104
80	0.3	0.102	0.099	0.098	0.099	0.098

(continued on next page)

n	$\bar{\pi}$	k=2	k=3	k=4	k=5	k=6
85	0.3	0.099	0.099	0.101	0.102	0.100
90	0.3	0.098	0.100	0.099	0.097	0.099
100	0.3	0.101	0.100	0.099	0.100	0.102
110	0.3	0.099	0.100	0.100	0.100	0.103
130	0.3	0.100	0.101	0.102	0.098	0.098
150	0.3	0.099	0.100	0.099	0.100	0.101
170	0.3	0.101	0.100	0.100	0.101	0.101
190	0.3	0.101	0.100	0.100	0.100	0.099
70	0.4	0.101	0.104	0.094	0.094	0.099
75	0.4	0.106	0.096	0.108	0.107	0.101
80	0.4	0.094	0.107	0.094	0.097	0.097
85	0.4	0.100	0.095	0.104	0.107	0.101
90	0.4	0.106	0.100	0.097	0.097	0.099
100	0.4	0.097	0.095	0.106	0.101	0.104
110	0.4	0.101	0.101	0.104	0.105	0.105
130	0.4	0.100	0.096	0.097	0.099	0.098
150	0.4	0.100	0.098	0.098	0.105	0.100
170	0.4	0.104	0.101	0.104	0.096	0.098
190	0.4	0.104	0.102	0.103	0.100	0.104
70	0.5	0.108	0.088	0.101	0.101	0.091
75	0.5	0.089	0.102	0.091	0.088	0.113
80	0.5	0.097	0.108	0.099	0.101	0.093
85	0.5	0.107	0.095	0.116	0.092	0.114
90	0.5	0.090	0.109	0.095	0.099	0.094
100	0.5	0.104	0.100	0.090	0.095	0.092
110	0.5	0.092	0.091	0.116	0.091	0.089
130	0.5	0.094	0.101	0.099	0.109	0.111
150	0.5	0.094	0.107	0.109	0.094	0.097
170	0.5	0.093	0.110	0.091	0.104	0.109
190	0.5	0.093	0.098	0.096	0.112	0.093

Table 4.2

The exact power of the test of the null hypothesis H_0 with k populations, sample size n, and nominal size $\alpha=0.10$, under the asymptotic least favourable configuration $\pi^*(\delta)$.

δ	n	k=2	k=3	k=4	k=5	k=6
0.10	70	0.339	0.242	0.234	0.214	0.185
0.10	75	0.320	0.278	0.228	0.203	0.227
0.10	80	0.348	0.299	0.250	0.232	0.205
0.10	85	0.382	0.290	0.288	0.225	0.244
0.10	90	0.364	0.325	0.263	0.246	0.221
0.10	95	0.387	0.301	0.298	0.283	0.257
0.10	100	0.418	0.333	0.274	0.257	0.233
0.10	110	0.421	0.339	0.340	0.267	0.244
0.10	130	0.477	0.401	0.348	0.336	0.315
0.10	150	0.524	0.456	0.407	0.342	0.322
0.10	170	0.566	0.503	0.408	0.397	0.379
0.10	190	0.607	0.519	0.455	0.446	0.379
0.20	70	0.782	0.654	0.618	0.572	0.516
0.20	75	0.779	0.708	0.626	0.575	0.589
0.20	80	0.813	0.741	0.670	0.629	0.579
0.20	85	0.846	0.748	0.723	0.637	0.644
0.20	90	0.843	0.790	0.713	0.677	0.633
0.20	95	0.866	0.782	0.759	0.728	0.689
0.20	100	0.890	0.817	0.751	0.718	0.679
0.20	110	0.904	0.841	0.824	0.753	0.718
0.20	130	0.943	0.903	0.864	0.846	0.822
0.20	150	0.966	0.941	0.916	0.879	0.860
0.20	170	0.979	0.964	0.934	0.924	0.912
0.20	190	0.988	0.974	0.959	0.953	0.930

Table 4.3

The exact powers of the test of the null hypothesis H_0 with $k=3$ populations, sample size n, and nominal size $\alpha=0.10$, at the asymptotic least favourable configuration

$\pi^*(\delta)=((1-\delta)/2,1/2,(1+\delta)/2)$ and at $\pi'(\delta)=((1-\delta)/2,(1-\delta)/2,(1+\delta)/2)$.

δ	n	π^*	π'
0.10	70	0.242	0.294
0.10	75	0.278	0.335
0.10	80	0.291	0.361
0.10	85	0.290	0.353
0.10	90	0.325	0.393
0.10	95	0.301	0.370
0.10	100	0.333	0.405
0.10	110	0.339	0.417
0.10	130	0.401	0.490
0.10	150	0.456	0.552
0.10	170	0.503	0.605
0.10	190	0.519	0.628
0.20	70	0.652	0.769
0.20	75	0.708	0.815
0.20	80	0.741	0.845
0.20	85	0.748	0.851
0.20	90	0.790	0.883
0.20	95	0.782	0.880
0.20	100	0.817	0.905
0.20	110	0.841	0.922
0.20	130	0.903	0.961
0.20	150	0.941	0.980
0.20	170	0.964	0.990
0.20	190	0.974	0.994

Table 4.4

Comparison of the exact powers at the asymptotic least favourable configuration $\pi^*(\delta)=((1-\delta)/2, 1/2, (1+\delta)/2)$ and the power using normal approximation (4.2), of the test of the null hypothesis H_0 with $k=3$ populations, sample size n, and and nominal size $\alpha=0.10$.

δ	n	exact power	approximate power
0.10	70	0.242	0.264
0.10	75	0.278	0.276
0.10	80	0.291	0.288
0.10	85	0.290	0.299
0.10	90	0.325	0.311
0.10	95	0.301	0.322
0.10	100	0.333	0.334
0.10	110	0.339	0.356
0.10	130	0.401	0.400
0.10	150	0.456	0.443
0.10	170	0.503	0.484
0.10	190	0.519	0.523
0.20	70	0.652	0.678
0.20	75	0.708	0.706
0.20	80	0.741	0.732
0.20	85	0.748	0.756
0.20	90	0.790	0.778
0.20	95	0.782	0.799
0.20	100	0.817	0.817
0.20	110	0.841	0.851
0.20	130	0.903	0.902
0.20	150	0.941	0.936
0.20	170	0.964	0.960
0.20	190	0.974	0.975

Chapter 5

Multiple comparisons with a control

5.1 Multiple comparisons with a control for normal data

In this chapter we consider the problem of testing k treatments with a control. In the first section, we consider this problem under the usual assumptions that observations from the control are independently normally distributed with mean μ_0 and variance σ^2, and that observations from the ith treatment are independently normally distributed with mean μ_i and variance σ^2, $1 \leq i \leq k$. The common variance σ^2, and the population means $\mu_0, \mu_1, \ldots, \mu_k$ are all unknown, and it is assumed that the experimenter is interested in making inferences on the k differences $\mu_i - \mu_0$, $1 \leq i \leq k$. In the next section, we consider the similar problem for Bernoulli data.

A common statistical problem is that of comparing simultaneously k (≥ 2) treatments with a control. The usual model employed for this problem is that the experimenter has n_0 independent normally distributed observations from the control

$$X_{0j} \sim N(\mu_0, \sigma^2), \qquad 1 \leq j \leq n_0$$

and n independent normally distributed observations from each of the k treatments

$$X_{ij} \sim N(\mu_i, \sigma^2), \qquad 1 \leq j \leq n,\ 1 \leq i \leq k.$$

The mean of the control population μ_0 and the k treatment means μ_i, $1 \leq i \leq k$, are unknown, and the experimenter is interested in making inferences on the k differences $\mu_i-\mu_0$, $1\le i\le k$. The error variance σ^2 is assumed to be common to the control and the k treatments, and is unknown.

Dunnett (1955,1964) showed how to construct a simultaneous set of confidence intervals (one-sided or two-sided) for these differences $\mu_i-\mu_0$, $1\le i\le k$ as follows. Define the sample means

$$\overline{X}_0 = \sum_{j=1}^{n_0} X_{0j}/n_0 \sim N(\mu_0, \sigma^2/n_0),$$

and

$$\overline{X}_i = \sum_{j=1}^{n} X_{ij}/n \sim N(\mu_i, \sigma^2/n), \qquad 1\le i\le k,$$

and let S^2 be distributed independently of these sample means as a $\sigma^2\chi^2_\nu/\nu$ random variable for some degrees of freedom ν. Usually, S^2 will be obtained from the following expression with $\nu = n_0+kn-(k+1)$:

$$S^2 = \frac{\sum_{j=1}^{n_0}(X_{0j}-\overline{X}_0)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2}{n_0+kn-(k+1)}.$$

Dunnett's procedures use the following two test statistics

$$T_1 = \max_{1\le i\le k}\left[\frac{\overline{X}_i-\overline{X}_0}{S\sqrt{\frac{1}{n_0}+\frac{1}{n}}}\right] \quad\text{and}\quad T_2 = \max_{1\le i\le k}\left[\frac{|\overline{X}_i-\overline{X}_0|}{S\sqrt{\frac{1}{n_0}+\frac{1}{n}}}\right]$$

for constructing one-sided and two-sided confidence intervals respectively. Under the condition $\mu_1-\mu_0 = \ldots = \mu_k-\mu_0 = 0$, the statistic T_1 has a k-dimensional t-distribution with ν degrees of freedom and equal correlations $\rho = \dfrac{n}{n+n_0}$, while T_2 has a similar $|t|$-distribution. Let the upper α point of these two distributions be labelled $T^{\alpha}_{k,\nu,\rho}$ and $|T|^{\alpha}_{k,\nu,\rho}$ respectively. One-sided and two-sided size α tests of the null hypothesis H_0 may then be made using the statistics T_1 and T_2 respectively, namely,

$$\text{reject } H_0 : \mu_i-\mu_0 \le 0,\ 1\le i\le k \quad \text{iff} \quad T_1 > T^{\alpha}_{k,\nu,\rho} \tag{5.1}$$

and

$$\text{reject } H_0 : \mu_i-\mu_0 = 0,\ 1\le i\le k \quad \text{iff} \quad T_2 > |T|^{\alpha}_{k,\nu,\rho}. \tag{5.2}$$

These tests produce the following sets of $1-\alpha$ level simultaneous confidence intervals. The set of simultaneous one-sided confidence intervals is

$$\mu_i-\mu_0 \ \geq \ \bar{X}_i-\bar{X}_0 - T^{\alpha}_{k,\nu,\rho} S\sqrt{\frac{1}{n_0}+\frac{1}{n}}\,, \qquad 1\leq i\leq k,$$

and the set of simultaneous two-sided confidence intervals is

$$\mu_i-\mu_0 \ \in \ \bar{X}_i-\bar{X}_0 \pm |T|^{\alpha}_{k,\nu,\rho} S\sqrt{\frac{1}{n_0}+\frac{1}{n}}\,, \qquad 1\leq i\leq k.$$

Both sets of confidence intervals have a joint confidence level of exactly $1-\alpha$. A test will reject the null hypothesis H_0 iff the respective set of confidence intervals includes a confidence interval which does not contain zero. When σ^2 is assumed to be known, we only need to put $\nu = \infty$, thus $S = \sigma$ which is assumed to be known, in the above procedures. Some tables of the critical points $T^{\alpha}_{k,\nu,\rho}$ and $|T|^{\alpha}_{k,\nu,\rho}$ may be found, for example, in Tamhane & Hochberg (1987), and Bechhofer and Dunnett (1988). Bechhofer (1969), Bechhofer & Nocturne (1972) and Bechhofer & Tamhane (1983) have shown how to choose the sample sizes to guarantee confidence intervals of a certain predetermined length when the error variance σ^2 is assumed known. However, we propose that the sample sizes should be chosen to directly satisfy certain prespecified conditions on the power function of the test being used. In the following, we consider the one-sided situation first.

If the experimenter is interested only in determining whether any of the the k treatments is better than the control, i.e. whether $\mu_i - \mu_0 > 0$ for some i, then it is appropriate to perform a one-sided test and to construct a set of simultaneous one-sided confidence intervals. For this problem we define two subsets of the parameter space $\mathbf{R}^{k+1}$ for $\mu = (\mu_0,\mu_1,\ldots,\mu_k)$, namely the "null set"

$$\Theta_{01} = \{ \ \mu : \mu_i-\mu_0 \leq 0, \quad 1\leq i\leq k \ \},$$

and the difference zone

$$\Theta_{A1} = \{ \ \mu : \mu_i - \mu_0 \geq a\sigma, \ \text{for at least one} \ i, \ 1 \leq i \leq k \ \}.$$

We are interested in testing the null hypothesis H_0: $\mu \in \Theta_{01}$, and shall reject the null hypothesis iff $T_1 > b$ for some critical point b. We stipulate the following probability requirements:

$$\mu \in \Theta_{01} \quad \Rightarrow \quad P(\text{ reject } H_0 \ ) \leq \alpha \tag{5.3}$$

and

$$\mu \in \Theta_{A1} \quad \Rightarrow \quad P(\text{ reject } H_0) \geq \beta. \tag{5.4}$$

These probability requirements have simple intuitive appeal and are easily interpretable. The experimenter can choose α, β and a to give the experiment any desired amount of sensitivity. For the given values of α, β, a and k predetermined by the experimenter, our goal is then to choose the sample sizes n_0 and n to satisfy the probability requirements (5.3) and (5.4), and to minimize the total sample size $N = n_0 + kn$ (assuming that the costs of sampling from the treatments and the control are equal).

To achieve this goal we first obtain an expression for the probability of rejecting the null hypothesis:

$$1 - P(\text{ reject } H_0) = P(\ T_1 \leq b\ )$$

$$= P\left[\frac{\bar{X}_i - \bar{X}_0}{S\sqrt{\frac{1}{n_0}+\frac{1}{n}}} \leq b;\ 1 \leq i \leq k \right]$$

$$= \int_{s=0}^{\infty} f(s)\ W_1(s)\ ds, \tag{5.5}$$

where $f(s)$ is the p.d.f. of a χ_ν^2/ν random variable, and

$$W_1(s) = P\left[\bar{X}_i - \bar{X}_0 \leq b\sqrt{s}\,\sigma\sqrt{\frac{1}{n_0}+\frac{1}{n}};\ 1 \leq i \leq k \right]$$

$$= \int_{x=-\infty}^{\infty} \varphi(x) \prod_{i=1}^{k} \Phi\left[\frac{-(\mu_i-\mu_0)\sqrt{n}}{\sigma} + x\sqrt{\frac{n}{n_0}} + b\sqrt{s}\sqrt{1+\frac{n}{n_0}} \right] dx,$$

where $\varphi(x)$ and $\Phi(x)$ are respectively the p.d.f. and c.d.f. of a standard normal random variable. It is clear from expression (5.5) that $P(\text{ reject } H_0)$ is increasing in each of the differences $\mu_i - \mu_0$, $1 \leq i \leq k$. Therefore, in order to satisfy condition (5.3) it is sufficient to choose the critical point b so that

$$P(\ T_1 > b \mid \mu_1-\mu_0 = \ldots = \mu_k-\mu_0 = 0\ ) = \alpha,$$

and hence, it follows from the discussion in the introduction that $b = T^{\alpha}_{k,\nu,\rho}$, with $\rho = n/(n+n_0)$.

In addition, if we define the least favourable configuration of the mean vectors μ to be those vectors μ in the difference zone Θ_{A1} which minimize the probability of rejecting the null hypothesis H_0, then clearly these least favourable configurations are of the form

$$\mu_1-\mu_0 = \dots = \mu_{k-1}-\mu_0 = -\infty, \qquad \mu_k-\mu_0 = a\sigma$$

or some permutation of this among the indices i, $1 \le i \le k$. Under this least favourable configuration, we have

$$P(\text{ reject } H_0 \text{ }) = P\left[\frac{N(\, a/\sqrt{\frac{1}{n_0}+\frac{1}{n}}\, , 1\,)}{\sqrt{\chi_\nu^2/\nu}} \ge T_{k,\nu,\rho}^{\alpha}\right] \tag{5.6}$$

where $\nu = n_0+kn-k-1$. Therefore we have to choose the sample sizes n_0 and n to minimize the total sample size N under the condition that expression (5.6) is no less than β. It is useful now to consider the sample sizes n_0 and n as continuous quantities. Notice that

$$n = \frac{N\rho}{1-\rho+k\rho}\ , \ \text{and} \quad n_0 = \frac{N(1-\rho)}{1-\rho+k\rho} \tag{5.7}$$

so that expression (5.6) may be rewritten as

$$P\left[\frac{N(\, a\sqrt{N}\sqrt{\frac{\rho(1-\rho)}{1+(k-1)\rho}}\, , 1\,)}{\sqrt{\chi_\nu^2/\nu}} \ge T_{k,\nu,\rho}^{\alpha}\right]. \tag{5.8}$$

An important question is how, for a fixed total sample size N, should the correlation ρ be chosen so as to maximize expression (5.8). In other words, for a fixed total sample size N, how should observations be allocated between the control and the treatments in order to maximize the least favourable power of the difference zone Θ_{A1}. It is not possible to provide a simple exact answer to this question due to the dependence of the critical point $T_{k,\nu,\rho}^{\alpha}$ on the correlation ρ. Nevertheless, investigation shows that $T_{k,\nu,\rho}^{\alpha}$ changes only very slightly as ρ changes. Therefore, it is sensible to choose the correlation ρ so as to maximize $\frac{\rho(1-\rho)}{1+(k-1)\rho}$, which is achieved at

$$\hat{\rho} = \frac{1}{1+\sqrt{k}}\ .$$

This choice of the correlation ρ maximizes expression (5.8), ignoring small changes in the critical point. The correlation $\hat{\rho}$ produces a sampling ratio of $n_0 = n\sqrt{k}$, which has previously been suggested as a good sampling ratio.

It is now necessary to find the smallest value of N so that expression (5.8) with $\rho = \hat{\rho}$ is greater than or equal to β, i.e.

$$P\left[\frac{N\left(\frac{a\sqrt{N}}{1+\sqrt{k}},\ 1\right)}{\sqrt{\chi^2_\nu/\nu}} \geq T^{\alpha}_{k,\nu,\hat{\rho}}\right] \geq \beta \tag{5.9}$$

where $\nu = N - (k+1)$. A simple computer program will now evaluate expression (5.9) and allow the required value of N to be determined. A slight snag is that the critical points $T^{\alpha}_{k,\nu,\rho}$ are only tabulated for certain values of ρ and ν. Nevertheless, they are easy to calculate, and it is simple to incorporate their calculation into the computer program. Also, since $T^{\alpha}_{k,\nu,\rho}$ is known to be decreasing in both ρ and ν, existing tables may be used to provide "conservative" values of the critical points needed in equation (5.9) which will ensure that the probability requirements are met.

Let the smallest value of N satisfying equation (5.9) be denoted by $\hat{N}$. Then the required sample sizes $\hat{n}_0$ and $\hat{n}$ may be determined from equations (5.7) evaluated at $\hat{N}$ and $\hat{\rho}$, i.e.

$$\hat{n} = \frac{\hat{N}}{\sqrt{k}+k}, \quad \text{and} \quad \hat{n}_0 = \frac{\hat{N}}{1+\sqrt{k}}.$$

These will, of course, not in general be integer values. The probability requirements will be met if they are both rounded up to the next integer. In addition, it may be possible to satisfy the probability requirements by rounding one of the numbers up and one of the numbers down, thus saving costs. Again, the computer program may be employed to discover whether this is possible. A useful rule in practice, it seems, is to round up the treatment sample size n, and then choose the control sample size n_0 to be the smallest integer such that $n_0+kn \geq \hat{N}$.

Some illustrative examples of the required sample sizes calculated using this procedure are contained in Table 5.1 for a size of $\alpha=0.05$, and the difference zone specified with $a=1.0$. For $k=2, \ldots, 6$, and $\beta=0.95$, 0.90, 0.80, 0.70, and 0.60, the table contains the required sample sizes n_0 and n, the total number of observation required N, and the critical point $T^{0.05}_{k,\nu,\rho}$. The precision was set at a level sufficient to guarantee agreement with independent calculations (e.g. previously published tables of critical points).

An interesting phenomenon which can be observed from Table 5.1 is that for a fixed value of the power level β, the treatment sample size n seems to be effectively independent of the number of treatments k. An advantage of this is that it would allow the experimenter to add additional treatments to the experiment at a late stage and maintain the probability level required without resampling from the previous treatments.

It may well be the case that, for certain experiments the experimenter will know that the differences $\mu_i - \mu_0$ are all larger than some negative constant. In other words, the experimenter may be able to state a bound $-L$, $0 < L < \infty$, and be confident that $\mu_i - \mu_0 \geq -L\sigma$, $0 \leq i \leq k$. In such a situation, the least favourable configuration

$$\mu_1 - \mu_0 = \ldots = \mu_{k-1} - \mu_0 = \infty, \quad \mu_k - \mu_0 = a\sigma$$

is not realistic, and guaranteeing the power level defined in (5.6) will be unnecessarily conservative. A natural change for this situation is to define the difference zone $\Theta_{A1}(L)$ by

$$\Theta_{A1}(L) = \{\ \mu :\ \max_{1 \leq i \leq k} (\mu_i - \mu_0) \geq a\sigma,\ \min_{1 \leq i \leq k} (\mu_i - \mu_0) \geq -L\sigma\ \}$$

for some positive constant a, L, and the power requirement is then

$$\mu \in \Theta_{A1}(L) \qquad \Rightarrow \qquad P(\text{reject } H_0) \geq \beta$$

for a fixed power level β. It is clear from expression (5.5) that the least favourable configurations for the difference zone $\Theta_{A1}(L)$ are of the form

$$\mu_1 - \mu_0 = \ldots = \mu_{k-1} - \mu_0 = -L\sigma, \quad \mu_k - \mu_0 = a\sigma\ .$$

Under this least favourable configuration, we have, in expression (5.5),

$$W_1(s) = \int_{x=-\infty}^{\infty} \varphi(x) \prod_{i=1}^{k} \Phi^{k-1} \left[L\sqrt{n} + x\sqrt{\frac{n}{n_0}} + b\sqrt{s}\sqrt{1+\frac{n}{n_0}} \right]$$
$$\times\ \Phi \left[-a\sqrt{n} + x\sqrt{\frac{n}{n_0}} + b\sqrt{s}\sqrt{1+\frac{n}{n_0}} \right] dx,$$

where $b = T^{\alpha}_{k,\nu,\rho}$. Again, we have to choose the sample sizes n_0 and n to minimize the total sample size N under the condition that the least favourable power for the difference zone $\Theta_{A1}(L)$ is no less than β. This problem can be treated in the similar way as in the preceding discussion, the details are omitted here. Next we consider the two-sidered test.

If the experimenter is interested in determining whether any of the k treatments are different to the control, i.e. whether $\mu_i - \mu_0 \neq 0$ for some i, then it is appropriate to perform a two-sided test and to construct a set of two-sided confidence intervals. In this case we define the "null set"

$$\Theta_{02} = \{ \ \mu: \ \mu_i - \mu_0 = 0, \ \ 1 \le i \le k \ \},$$

and the difference zone

$$\Theta_{A2} = \{ \ \mu: \ | \ \mu_i - \mu_0 \ | \ge a\sigma \ \text{ for at least one } \ i, \ \ 1 \le i \le k \ \}.$$

Again we are interested in testing the null hypothesis

$$H_0 : \ \ \mu \in \Theta_{02},$$

and our probability requirements are

$$\mu \in \Theta_{02} \quad \Rightarrow \quad P(\text{ reject } \ H_0 \ ) \le \alpha \tag{5.10}$$

and

$$\mu \in \Theta_{A2} \quad \Rightarrow \quad P(\text{ reject } \ H_0 \ ) \ge \beta \ , \tag{5.11}$$

where the null hypothesis is rejected iff $T_2 > b$ for a suitable critical point b. This power approach has been independently proposed by Bristol(1989) in the context of designing clinical trials, in which the author gave a method to calculate a lower bound on $\min\limits_{\mu \in \Theta_{A2}} P(\text{ reject } \mathrm{H}_0 \ )$ when σ^2 is assumed known and $n_0 = n$. For given values of α, β, a and k predetermined by the experimenter, we have to find values of the sample sizes n_0 and n to satisfy the probability requirements (5.10) and (5.11), and yet to minimise the total sample size $N = n_0 + kn$.

As before, it is clear from the discussion in the introduction that if $|T|_{k,\nu,\rho}^{\alpha}$ is used as the critical point, then

$$\mu \in \Theta_{02} \quad \Rightarrow \quad P(\text{ reject } \ H_0 \ ) = \alpha,$$

so that the probability requirement (5.10) is satisfied. The probability of rejecting the null hypothesis H_0 may then be written as

$$P(\text{ reject } H_0 \ ) = 1 - \int_{s=0}^{\infty} f(s) \ W_2(s,N,\rho,\delta) \ ds \tag{5.12}$$

where, as before, $f(s)$ is the p.d.f. of a χ_ν^2/ν random variable, and

$$W_2(s,N,\rho,\delta) = P(\ | \ Y_i \ | \le \sqrt{s} \ |T|_{k,\nu,\rho}^{\alpha}; \ \ 1 \le i \le k \ ) \ , \tag{5.13}$$

where $\mathbf{Y} = (\ Y_1 , \ ..., \ Y_k \ )$ has a k-dimensional normal distribution with unit variances, covariances all equal to $\rho = n/(n+n_0)$, and mean vector

$$\left[\sigma\sqrt{\frac{1}{n}+\frac{1}{n_0}}\right]^{-1}\boldsymbol{\delta} = \sqrt{N}\sqrt{\frac{\rho(1-\rho)}{1+(k-1)\rho}}\,\frac{1}{\sigma}\,\boldsymbol{\delta}$$

where

$$\boldsymbol{\delta} = (\,\mu_1-\mu_0,\,\ldots,\,\mu_k-\mu_0\,).$$

In order to guarantee that the probability requirement (5.11) is satisfied, it is now desirable to determine the least favourable configuration of the vector $\boldsymbol{\mu}$ in the difference zone Θ_{A2}, i.e. the vectors $\boldsymbol{\mu}$ in Θ_{A2} which minimize the probability of rejecting H_0. Unfortunately, it is not possible, when the variance σ^2 is unknown, to determine the least favourable configuration or even to shrink Θ_{A2} to a useful subset which contains at least one least favourable configuration of Θ_{A2}. This makes numerical search of the least favourable power in Θ_{A2} complicated even for intermediate sized k. In the following, we first consider the known variance ($\nu=\infty$) situation. In this case, we can identify the form of the least favourable configurations of Θ_{A2} to be

$$\boldsymbol{\delta} = \boldsymbol{\delta}(u) = a\sigma\,(\,\lambda,\,\ldots,\,\lambda,\,1\,)$$

for some $\lambda,\ 0<\lambda<1$. This property can be used to simplify the numerical search for the least favourable power. For the unknown variance case, we will give a lower bound on the least favourable power.

If the variance σ^2 is known, the probability of rejecting the null hypothesis H_0 is simply equal to $1 - W_2(1,N,\rho,\boldsymbol{\delta})$, where $W_2(\,)$ is defined in (5.13) with $\nu=\infty$. The following lemma establishes some basic properties of $W_2(1,N,\rho,\boldsymbol{\delta})$ as a function of $\boldsymbol{\delta}$.

Lemma 5.1 For $W_2(1,N,\rho,\,\boldsymbol{\delta})$, we have

(i) $W_2(1,N,\rho,\,\boldsymbol{\delta}) = W_2(1,N,\rho,\,-\boldsymbol{\delta})$.

(ii) $W_2(1,N,\rho,\,\boldsymbol{\delta}) = W_2(1,N,\rho,\,\pi\boldsymbol{\delta})$, for all $\pi\in\Pi$.

(iii) $W_2(1,N,\rho,\,\boldsymbol{\delta}) \geq W_2(1,N,\rho,\,u\boldsymbol{\delta})$, $u\geq 1$.

(iv) $W_2(1,N,\rho,\,\boldsymbol{\delta}) \geq W_2(1,N,\rho,\,\boldsymbol{\delta}^*)$, if $\boldsymbol{\delta}^* \gg \boldsymbol{\delta}$.

Proof of Lemma 5.1

As the proof is exactly the same as the proof of Lemma 3.1 in chapter 3, the reader is referred to the proof of Lemma 3.1 #

The following lemma gives the form of the least favourable configuration of the test defined in (5.2) with $\nu=\infty$ for the difference zone Θ_{A2}.

Lemma 5.2 For $W_2(1,N,\rho,\ \delta)$, we have

$$\max_{\mu \in \Theta_{A2}} W_2(1,N,\rho,\ \delta) = \max_{\lambda \in [-1,\ 1]} W_2(1,N,\rho,\ \delta(\lambda))\ ,$$

where $\delta = (\mu_1-\mu_0,\ \ldots,\ \mu_k-\mu_0)$ and $\delta(\lambda) = a\sigma(\lambda,\ \ldots,\ \lambda,\ 1) \in \mathbf{R}^k$.

Proof of Lemma 5.2

As a consequence of property (iii) of Lemma 5.1, it is sufficient to restrict attention to the vectors μ for which $\max_{1\le i\le k} |\mu_i-\mu_0| = a\sigma$. Also, from the properties (i) and (ii) of Lemma 5.1 we can assume $-(\mu_k-\mu_0) \le \mu_1-\mu_0 \le \mu_2-\mu_0 \le \ldots \le \mu_k-\mu_0$ and $\mu_k-\mu_0 = a\sigma$ without loss of generality. Then as a consequence of the property (iv) of Lemma 5.1 and the fact

$$(\ \mu_1-\mu_0,\ \ldots\ ,\ \mu_{k-1}-\mu_0,\ \mu_k-\mu_0\ ) \gg a\sigma\ (\ \lambda,\ \ldots\ ,\ \lambda,\ 1\ )$$

where $\lambda = \dfrac{1}{a\sigma(k-1)} \sum_{i=1}^{k-1}(\mu_i-\mu_0) \in [0,\ 1]$, the lemma follows immediately #

In Lemma 5.4 below we ascertain that the actual value of λ which maximises $W_2(1,N,\rho,\delta(\lambda))$ is in $[0,\ 1]$. To prove this we first need another simple lemma.

Lemma 5.3 Let $f(x)$ and $g(x)$ be symmetric functions such that $x_1 \ge x_2 \ge 0$ implies $f(x_2) \ge f(x_1) \ge 0$ and $g(x_2) \ge g(x_1) \ge 0$. Also, let $h(x)$ be an odd function with $x\,h(x) \ge 0$. Then for constants $a,\ b \ge 0$ we have

$$\int_{-\infty}^{\infty} f(x)\ g(x-a)\ h(x+b)\ dx\ \ge\ 0.$$

Proof of Lemma 5.3

Notice that

$$\int_{-\infty}^{\infty} f(x)\ g(x-a)\ h(x+b)\ dx = \int_{-\infty}^{\infty} f(x-b)\ g(x-a-b)\ h(x)\ dx$$

$$= \int_{0}^{\infty} \left[f(x-b)\ g(x-a-b) - f(x+b)\ g(x+a+b) \right] h(x)\ dx\ \ge\ 0,$$

the final inequality following from the fact that for $x \ge 0$ we have

$$h(x) \ge 0, \qquad f(x-b) \ge f(x+b)$$

and

$$g(x-a-b) \ge g(x+a+b).$$

This completes the proof of lemma 5.3 #

We now state Lemma 5.4.

Lemma 5.4 For the $W_2(1,N,\rho,\delta)$ defined in equation (5.13) we have

$$\mu \in \Theta_{A2} \quad \Rightarrow \quad W_2(1,N,\rho,\delta) \le \max_{\lambda \in [0,\ 1]} W_2(1,N,\rho,\delta(\lambda))$$

Proof of Lemma 5.4

As a consequence of Lemma 5.2, in order to prove the lemma it is now sufficient to establish that

$$\max_{-1 \le \lambda \le 1} W_2(1,N,\rho,\delta(\lambda)) = \max_{0 \le \lambda \le 1} W_2(1,N,\rho,\delta(\lambda)). \tag{5.14}$$

Equation (5.14) will be established by showing that

$$\lambda \le 0 \quad \Rightarrow \quad \frac{d\, W_2(1,N,\rho,\delta(\lambda))}{d\lambda} \ge 0\,. \tag{5.15}$$

Notice that

$$W_2(1,N,\rho,\delta(\lambda)) = P\{\ |\ Z_i - Z_0\ |\ \le t_1;\ \ 1 \le i \le k\ \}\ ,$$

where

$$t_1 = |T|^{\alpha}_{k,\infty,\rho} \sqrt{\frac{1}{n} + \frac{1}{n_0}}\ ,$$

and where Z_0 and the Z_i, $1 \le i \le k$, are independent random variables with

$$Z_0 \sim N(\,0,\ \frac{1}{n_0}\,),$$

$$Z_i \sim N(\,\lambda a,\ \frac{1}{n}\,), \qquad 1 \le i \le k-1$$

$$Z_k \sim N(\,a,\ \frac{1}{n}\,).$$

Thus we can write

$$W_2(1,N,\rho,\delta(\lambda)\,) = \int_{x=-\infty}^{\infty} f_1(x)\; g_1(x-a)\;(\; g_1(x-\lambda a)\;)^{k-1}\; dx$$

where

$$f_1(x) = \sqrt{n_0}\;\varphi(\;\sqrt{n_0}\; x\;), \tag{5.16}$$

and

$$g_1(x) = \Phi(\;\sqrt{n}\;\; (x+t_1)\;) - \Phi(\;\sqrt{n}\;\; (x-t_1)\;), \tag{5.17}$$

and hence

$$\frac{d\; W_2(1,N,\rho,\delta(\lambda)\,)}{d\lambda} = \int_{x=-\infty}^{\infty} f_1(x)\; g_1(x-a)\; h_1(x-\lambda a)\; dx \tag{5.18}$$

where

$$h_1(x) = a\,\sqrt{n}\;\,(k-1)\;(\; g_1(x)\;)^{k-2}\;\Big[\varphi(\sqrt{n}\;\,(x-t_1)) - \varphi(\sqrt{n}\;\,(x+t_1))\Big]. \tag{5.19}$$

Now, we note from equations (5.16), (5.17) and (5.19) that the functions $f_1(x)$, $g_1(x)$ and $h_1(x)$ satisfy the conditions imposed in lemma 5.3 on the functions $f(x)$, $g(x)$ and $h(x)$ respectively, and hence that Lemma 5.3 ensures that expression (5.18) is non-negative for $\lambda \le 0$. Thus expression (5.15) is satisfied, and this completes the proof of the lemma #

In fact, it is also true that

$$\frac{d\; W_2(1,N,\rho,\delta(\lambda)\,)}{d\lambda} > 0, \quad \text{for} \quad \lambda = 0\;,$$

and

$$\frac{d\; W_2(1,N,\rho,\delta(\lambda)\,)}{d\lambda} < 0, \quad \text{for} \quad \lambda = 1\;,$$

which can be seen from expression (5.18). Hence, the value λ at the least favourable configuration is in fact in $(0,\; 1)$.

So the least favourable configuration of the test defined in (5.2) for the difference zone Θ_{A2}, assuming that σ is known, has the form

$$(\;\mu_1-\mu_0,\;\ldots\;,\;\mu_k-\mu_0\;) = a\sigma\;(\;\lambda,\;\ldots\;,\;\lambda,\;1\;),$$

for some $\lambda \in (0,\; 1)$. The actual value of λ at the least favourable configuration depends in a complicated fashion on the parameters N, ρ, k, α, a. Denote the probability of rejecting H_0 at $\delta(\lambda)$ by

$$P(\text{ reject } H_0\ )|_{\delta(\lambda)} = 1 - W_2(1,N,\rho,\delta(\lambda)\ ) \tag{5.20}$$

$$= h(\ N,\rho,\lambda\ ).$$

The problem now is to find the minimum value of the total sample size N such that

$$\max_{0\le\rho\le 1}\ \min_{0\le\lambda\le 1}\ h(\ N,\rho,\lambda\ ) \ \ge\ \beta.$$

The maximisation over the correlation ρ arises because for a given total sample size N the allocation of observations between the control and the treatments can be chosen to maximize the least favourable power in the difference zone Θ_{A2}, and the minimisation over λ occurs because the probability requirement (5.11) needs to be guaranteed at the (unknown) least favourable configuration in the difference zone Θ_{A2}.

We address first the problem of the maximisation over the correlation ρ. The dependence of the function $h(\ N,\rho,\lambda\ )$ on ρ occurs in three ways: first, through the critical point $|T|^{\alpha}_{k,\infty,\rho}$, secondly through the correlation of the Y_i, and thirdly through the scale factor

$$\sqrt{\frac{\rho\ (1-\rho)}{1 + (k-1)\ \rho}} \tag{5.21}$$

of the mean vector of the Y_i's. Furthermore, computational investigation shows that the effect on $h(\ N,\rho,\lambda\ )$ of changes in ρ in the first two cases is slight compared with the effect on $h(\ N,\rho,\lambda\ )$ of changes in ρ in the third case. Thus a sensible choice of ρ might be

$$\hat{\rho} = \frac{1}{1 + \sqrt{k}}$$

which maximizes expression (5.21). Such a choice of ρ maximises the function $h(\ N,\rho,\lambda\ )$ ignoring the first two dependences on ρ described above (this can been seen from the property (iii) of Lemma 5.1). Again, the correlation $\hat{\rho}$ produces a sample ratio of $n_0 = n\sqrt{k}$ which has previously been suggested as a good sampling ratio for generating simultaneous confidence intervals. In Table 5.2 we present some calculations which confirm that, for the cases considered, allocating observations according (as closely as possible) to the correlation $\hat{\rho}$ is in fact optimal. For k=3 with N=120 and N=76, we investigated the changes in the power under the least favourable configuration as the sampling ratio n_0/n

is varied while the total sample size N remains fixed. From the results given in Table 5.2, it can be seen that the power is largest when the sampling ratio is $n_0/n \approx \sqrt{3} \approx 1.73$. Other investigations for different sets of parameters also support this indication that there is nothing to be gained by using sampling ratios, n_0/n, which are not as close as possible to $\sqrt{k}$.

Since the correlation $\hat{\rho}$ is independent of λ, our problem is now to find the smallest total sample size N such that

$$\min_{0 \le \lambda \le 1} h(N, \hat{\rho}, \lambda) \ge \beta .$$

Numerical investigation shows that $h(N,\hat{\rho},\lambda)$ is unimodal as a function of $\lambda \in (0, 1)$. So, a simple computer program can be employed to find the required value $\hat{N}$. Notice that each time N is changed, a further search needs to be performed over λ to find the value $\hat{\lambda}(N)$ at which the minimum is achieved. The function $h(N,\hat{\rho},\lambda)$ can be evaluated using the expression (5.20) with

$$W_2(1,N,\hat{\rho},\lambda) = \int_{x=-\infty}^{+\infty} \varphi(x) \times$$

$$\Big\{ \Phi\left[x\, k^{-1/4} + \sqrt{1+1/\sqrt{k}}\, |T|^{\alpha}_{k,\infty,\hat{\rho}} - \lambda a \sqrt{N}\, (k+\sqrt{k})^{-1/2} \right]$$

$$- \Phi\left[x\, k^{-1/4} - \sqrt{1+1/\sqrt{k}}\, |T|^{\alpha}_{k,\infty,\hat{\rho}} - \lambda a \sqrt{N}\, (k+\sqrt{k})^{-1/2} \right] \Big\}^{k-1} \times$$

$$\Big\{ \Phi\left[x\, k^{-1/4} + \sqrt{1+1/\sqrt{k}}\, |T|^{\alpha}_{k,\infty,\hat{\rho}} - a \sqrt{N}\, (k+\sqrt{k})^{-1/2} \right]$$

$$- \Phi\left[x\, k^{-1/4} - \sqrt{1+1/\sqrt{k}}\, |T|^{\alpha}_{k,\infty,\hat{\rho}} - a \sqrt{N}\, (k+\sqrt{k})^{-1/2} \right] \Big\}\, dx.$$

The Fortran77 programs necessary to calculate the required critical values $|T|^{\alpha}_{k,\infty,\rho}$ and to find $\hat{N}$ have been developed.

Once the required value of the total sample size $\hat{N}$ has been found, the sample sizes $\hat{n}$ and $\hat{n}_0$ may be calculated from equations

$$\hat{n}_0 = \sqrt{k}\, \hat{n} \quad \text{and} \quad \hat{N} = \hat{n}_0 + k\, \hat{n},$$

which gives

$$\hat{n} = \frac{\hat{N}}{\sqrt{k} + k} \quad \text{and} \quad \hat{n}_0 = \frac{\hat{N}}{1 + \sqrt{k}}.$$

Again, these will, of course, not in general be integer values. A practical rule is to choose the treatment sample size n and the control sample size n_0 to be the integers such that $n_0 + kn = \hat{N}$, and n_0/n is as close to $\sqrt{k}$ as possible. Some illustrative examples of the required sample sizes are contained in Table 5.3. For $\alpha = 0.05$, $a = 1.0$, and $k = 2, \dots, 6$, the table contains the required sample sizes n_0 and n, the total number of observation required, N, for guaranteeing the power levels $\beta = 0.60, 0.70, 0.80, 0.90, 0.95$, and the critical point $|T|^{0.05}_{k,\infty,\rho}$. In addition, Table 5.3 contains, for completeness, values of $\hat{\lambda}$, the value of λ at the least favourable configuration in the difference zone Θ_{A2}, although this will not usually be of interest to the experimenter.

In the preceding discussion, the variance σ^2 is assumed to be known. If σ^2 is unknown, we will not be able to identify the form of the least favourable configuration of Dunnett's test procedure for the difference zone Θ_{A2}. So, the numerical search for the least favourable power is complicated even for k=4, for instance. However, as a first approximation, we can assume that σ^2 is known and equal to the value estimated from the available data, then follow the preceding discussion. Following, we obtain a lower bound on the least favourable power in Θ_{A2}. We can see from expression (5.12) that

$$\min_{\mu \in \Theta_{A2}} P(\text{ reject } H_0 \text{ }) = 1 - \max_{\mu \in \Theta_{A2}} \int_{s=0}^{\infty} f(s)\, W_2(s,N,\rho,\delta)\, ds$$

$$\geq 1 - \int_{s=0}^{\infty} f(s) \max_{\mu \in \Theta_{A2}} W_2(s,N,\rho,\delta)\, ds$$

$$= 1 - \int_{s=0}^{\infty} f(s) \max_{\lambda \in [0,\, 1]} W_2(s,N,\rho,\delta(\lambda))\, ds. \quad (5.22)$$

Expression (5.22) can be used as a lower bound on the least favourable power in the difference zone Θ_{A2}. The complexity of its numerical evaluation is independent of the number of treatments k, and it is less than a three dimensional integral. Table 5.4 contains this lower bounds for $\nu = 20$(top) and the least favourable powers for $\nu = \infty$(bottom), for different values of $h = \sqrt{N/(k+1)}\, a$ under the square-root sample allocation, while $\alpha = 0.1$, 0.05, k=2 to 8. From the entries of the table, we can see that the lower bounds for $\nu = 20$ can be lower than least favourable power for $\nu = \infty$ by as much as 0.101. If we search for the least favourable powers directly for $\nu = 20$, $k = 2$, 3, by using the NAG routines, then the results are very close to the lower bounds given in Table 5.4 (the first two columns). In fact, the largest difference is 0.001. This indicates that the lower bounds given by expression (5.22), with ν as large as 20, are good approximations for the least favourable powers of Θ_{A2}. In order to guarantee the least favourable power to be no less than β, we only need to guarantee the lower bound given in (5.22) to be no less than β.

In the next section we will consider the problem of comparisons with a control for Bernoulli data.

5.2 Multiple comparisons with a control for Bernoulli data

In this section the problem of testing the equality of several treatment Bernoulli probabilities and a control Bernoulli probability is considered, and attention is specially directed to forming an assessment of power properties of such a test.

Suppose that we have k independent treatments with Bernoulli responses in terms of their success probabilities $\pi_1, \ldots, \pi_k$, and that our data consist of n independent observations on each of the k treatments. Let Y_i denotes the number of successes from the ith treatment, the sample proportion of successes being $\hat{\pi}_i = Y_i/n$ $(1 \le i \le k)$. In addition, we assume that we have n_0 independent observations from the control population which are identically distributed as Bernoulli random variables with success probability π_0, and we denote the number of successes by Y_0, so that our estimate of π_0 is $\hat{\pi}_0 = Y_0/n_0$. In this situation we consider testing the null hypothesis

$$H_0 \; : \; \pi_1 - \pi_0 = \ldots = \pi_k - \pi_0 = 0$$

at size α, and stipulate that the power requirement is no less than β whenever

$$\max_{1 \le i \le k} | \; \pi_0 - \pi_i \; | \ge a$$

for some given value $a \in (0, 1)$. By using the acrsin-root transformation

$$\arcsin\sqrt{\hat{\pi}_i} \quad \overset{asy.}{\sim} \quad N(\arcsin\sqrt{\pi_i}, \; \frac{1}{4n}) , \qquad 0 \le i \le k \tag{5.23}$$

applied to the treatments and to the control, we derive a large sample test to meet the probability requirements, and show how to determine the necessary sample size.

If the experimenter is interested in determining whether any of the k treatments are different from the control population, i.e. whether $\pi_i - \pi_0 \neq 0$ for some i, then it is appropriate to test the following null hypothesis

$$H_0 : \pi_1 - \pi_0 = \ldots = \pi_k - \pi_0 = 0 . \tag{5.24}$$

Let the set Θ_A be defined by

$$\Theta_A = \{ \pi_+ = (\pi_0, \pi_1, \ldots, \pi_k) : \max_{1 \le i \le k} |\pi_i - \pi_0| \ge a \} \subset [0, 1]^{k+1}$$

for some fixed positive constant $a \in (0, 1)$. The problem which we address in this section is to find a size α test of the null hypothesis H_0 which satisfies the following power requirement

$$\pi_+ \in \Theta_A \quad \Rightarrow \quad P(\text{ reject } H_0) \ge \beta \tag{5.25}$$

for a given fixed power level β. Again this power requirement is readily interpretable by the experimenter, and the choice of the parameters α, β, and a allow the experimenter to achieve any desired level of sensitivity for the experiment.

We propose the following test procedure which asymptotically satisfies the power requirements. We define

$$T = \max_{1 \le i \le k} \frac{| \arcsin\sqrt{\hat{\pi}_i} - \arcsin\sqrt{\hat{\pi}_0} |}{\frac{1}{2}\sqrt{\frac{1}{n_0} + \frac{1}{n}}} ,$$

and we shall reject the null hypothesis H_0 iff

$$T > |T|^{\alpha}_{k,\infty,\rho} ,$$

where $|T|^{\alpha}_{k,\infty,\rho}$ is the upper α-point of the $|t|$-distribution with parameters k, ∞, ρ (see Tamhane & Hochberg (1987)), and where $\rho = \frac{n}{n+n_0}$. As a consequence of the asymptotic properties of the arcsin-root transformation given in equation (5.23), this test procedure is asymptotically of size α. Notice, however, that for fixed sample sizes n and n_0, the distribution of T under the null hypothesis H_0 depends on the particular values taken by the probabilities π_i. Also, in practice, it seems that for small sample sizes, a variance stabilizing correction of Anscombe (1948) is helpful in controlling the exact size of the test procedure, so that it is better to use $\hat{\pi}_i = (Y_i+3/8)/(n+3/4)$ ($1 \le i \le k$), and $\hat{\pi}_0 = (Y_0+3/8)/(n_0+3/4)$ in the test statistic rather than the actual exact sample proportions, and also $n+1/2$, $n_0+1/2$ in place of n, n_0.

We now address the problem of determining the sample sizes n_0 and n to satisfy the stipulated probability requirement (5.25) and to minimise the total sample size $N = n_0 + kn$. Define

$$h(\pi_+, n, n_0) = P\{ T > |T|^{\alpha}_{k,\infty,\rho} \mid \pi_+, n, n_0 \}$$

to be the probability of rejecting the null hypothesis H_0 for given values of π_+, n, and n_0. The asymptotic least favourable configurations of π_+ are defined to be those values of π_+ which are in the set Θ_A and which asymptotically minimise the function $h(\pi_+, n, n_0)$. The following lemma categorises these least favourable configurations.

Lemma 5.5 Define $\pi_+^*(u) = ((1-a)/2, u, \ldots, u, (1+a)/2)$. Then as $n, n_0 \to \infty$ we have

$$\pi_+ \in \Theta_A \quad \Rightarrow \quad h(\pi_+, n, n_0) \geq \min_{(1-a)/2<u<(1+a)/2} h(\pi_+^*(u), n, n_0).$$

Proof of Lemma 5.5

Without loss of generality assume that $\pi_1 \leq \ldots \leq \pi_k$ and $\pi_0 \leq (\pi_1 + \pi_k)/2$. Notice that

$$\pi_k - \pi_0 = \tilde{a} \quad \Rightarrow \quad \arcsin\sqrt{\pi_k} - \arcsin\sqrt{\pi_0} \geq b$$

where

$$b = \arcsin\sqrt{(1+\tilde{a})/2} - \arcsin\sqrt{(1-\tilde{a})/2} \; ,$$

with equality achieved only when

$$\pi_0 = \frac{1-\tilde{a}}{2}, \quad \pi_k = \frac{1+\tilde{a}}{2} \; .$$

In addition, it follows from Lemma 5.4, assuming the asymptotic normal distribution (5.23), that under the condition

$$\max_{1 \leq i \leq k} | \arcsin\sqrt{\pi_i} - \arcsin\sqrt{\pi_0} | \geq b \; ,$$

the power function $h(\pi_+, n, n_0)$ is minimised when

$$\arcsin\sqrt{\pi_0} = c$$

$$\arcsin\sqrt{\pi_i} = c + u^* \qquad 1 \leq i \leq k-1$$

$$\arcsin\sqrt{\pi_k} = c + b$$

for any constant c and some particular value of u^*, $0 < u^* < b$, depending on the parameters b, n, n_0, α and k. Furthermore, the power function in this case is increasing in b. Consequently, under the condition

$$\max_{1 \le i \le k} \mid \pi_i - \pi_0 \mid \ge a ,$$

the power function $h(\pi_+, n, n_0)$ is minimised when

$$\pi_0 = \frac{1-a}{2} , \qquad \pi_k = \frac{1+a}{2} ,$$

and for $1 \le i \le k-1$,

$$\pi_i = u \in (\frac{1-a}{2}, \frac{1+a}{2}) .$$

The proof is thus completed #

For a given total sample size N, the experimenter can choose how to allocate observations between the treatments and the control. According to the results of Section 5.1, it is optimal to allocate observations between the treatments and control in such a way that

$$n = n^*(N) = \frac{N}{\sqrt{k}+k}$$

and (5.26)

$$n_0 = n_0^*(N) = \frac{N}{1+\sqrt{k}} .$$

Such an allocation is optimal in the sense of maximising the least favourable power for a fixed total sample size N assuming the asymptotical normality of (5.23). Of course, (5.26) will not give integer values in general. A useful rule in practice, it seems, is to round the treatment sample size n to the nearest integer, and then choose the control sample size n_0 to be the integer such that $n_0 + kn = N$. Thus, in order to satisfy the power requirement (5.25) asymptotically, the total sample size N must be chosen so that

$$\min_{(1-a)/2<u<(1+a)/2} h(\, \pi_+^*(u), n^*(N), n_0^*(N) \,) \ge \beta . \tag{5.27}$$

A simple computer program may be employed to find such N using the following expression for the power function $h(\pi_+,n,n_0)$, assuming that the variance stabilizing correction is applied. Define

$$d = \frac{1}{2}\sqrt{\frac{1}{n_0+1/2} + \frac{1}{n+1/2}}\ |T|^{\alpha}_{k,\infty,\rho} .$$

Then

$$1 - h(\pi_+,n,n_0) =$$

$$P\{ \max_{1\le i\le k} | \arcsin\sqrt{\frac{Y_i+3/8}{n+3/4}} - \arcsin\sqrt{\frac{Y_0+3/8}{n_0+3/4}} | \le d \}$$

$$= \sum_{y=0}^{n_0} P(Y_0=y) \prod_{i=1}^{k} P\{ |\arcsin\sqrt{\frac{Y_i+3/8}{n+3/4}} - \arcsin\sqrt{\frac{y+3/8}{n_0+3/4}}| \le d \}$$

$$= \sum_{y=0}^{n_0} P(Y_0 = y) \prod_{i=1}^{k} P\{ V(y) \le Y_i \le U(y) \} , \tag{5.28}$$

where

$$V(y) = (n+3/4)\left\{ sin[max(0, -d + \arcsin\sqrt{\frac{y+3/8}{n_0+3/4}})]\right\}^2 - 3/8 ,$$

$$U(y) = (n+3/4)\left\{ sin[min(\frac{\pi}{2}, d + \arcsin\sqrt{\frac{y+3/8}{n_0+3/4}})]\right\}^2 - 3/8 .$$

The complexity of the exact evaluation of the power function $h(\pi_+^*(u),n,n_0)$ is independent of k (since $\pi_1 = \dots = \pi_{k-1} = u$ and so $k-1$ of the product terms in equation (5.28) are identical) but is increasing in the sample sizes n and n_0. In fact

$$1 - h(\pi_+^*(u),n,n_0) =$$

$$\sum_{y=0}^{n_0} P(Y_0 = y) P\{ V(y) \le Y_1 \le U(y) \}^{k-1} P\{ V(y) \le Y_k \le U(y) \} ,$$

where $Y_0 \sim B(n_0, (1-a)/2)$, $Y_1 \sim B(n, u)$ and $Y_k \sim B(n, (1+a)/2)$. The exact size of the test procedure when $\pi_1 = \dots = \pi_k = \pi_0 = \bar{\pi}$ may also be evaluated using expression (5.28), which becomes

$$1 - h(\bar{\pi}_+,n,n_0) =$$

$$\sum_{y=0}^{n_0} P(Y_0 = y) P\{ V(y) \le Y_1 \le U(y) \}^{k} ,$$

where $\bar{\pi}_+ = (\bar{\pi}, \ldots , \bar{\pi}) \in \mathbf{R}^{k+1}$, $Y_0 \sim B(n_0, \bar{\pi})$, and $Y_1 \sim B(n, \bar{\pi})$.

In Table 5.5 we present some calculations of the exact size of the test procedure for a nominal size $\alpha = 0.10$, for $k = 2, \ldots , 6$ populations, $\pi_i = \pi_0 = \bar{\pi} = 0.1, 0.2, \ldots , 0.5$, $1 \le i \le k$, and various sets of sample sizes $N (n, n_0)$ according to the optimal allocation formula (5.26) with n rounded to the nearest integer. These values were calculated using the variance stabilising correction suggested above. It can be seen from Table 5.5 that the convergence of exact size to the nominal size is not monotonic in the sample sizes due to the discrete nature of the problem.

In Table 5.6 we present some calculations of the exact value

$$\min_{(1-a)/2<u<(1+a)/2} h(\pi_+^*(u),n,n_0) \equiv \beta^* \tag{5.29}$$

for $a = 0.1, 0.2$, nominal size $\alpha = 0.10$, $k = 2, \ldots , 6$, and the sets of sample size $N (n, n_0)$ used in Table 5.5. It can be seen from Table 5.6 that in order to obtain power level $\beta \ge 0.90$ for $a = 0.2$ say, then quite large sample sizes are required, e.g. $n = 109$, $n_0 = 188$ for $k = 3$. For sample sizes of this order it would appear from Table 5.5 that the exact size is generally well approximated by the nominal size.

It may well be the case that, for certain experiments the experimenter will know that the probabilities π_i are all large or all small. In other words, the experimenter may be able to state a bound L, $1/2 \le L \le 1$, and be confident that $\pi_i \ge L$, $0 \le i \le k$, (or equivalently, $\pi_i \le 1-L$, $0 \le i \le k$). In such a situation, the asymptotic least favourable configuration $\pi_+^*(u)$ is not realistic, and guaranteeing the power level β^* defined in (5.29) will be unnecessarily conservative. A natural change for this situation is to define the difference zone $\Theta_A(a,L)$ by

$$\Theta_A(a,L) = \{ \pi_+=(\pi_0, \pi_1, \ldots , \pi_k) : \max_{1 \le i \le k} |\pi_i - \pi_0| \ge a, \min_{0 \le i \le k} \pi_i \ge L \}$$

for some positive constant a, and the power requirement is then

$$\pi_+ \in \Theta_A(a,L) \quad \Rightarrow \quad P(\text{ reject } H_0) \ge \beta$$

for a fixed power level β. This problem can be treated in the similar way as in the preceding discussion, the details are omitted here.

Table 5.1

Table of the required sample sizes n_0 from the control and n from the treatments for the one-sided test procedure with k treatments, $\alpha=0.05$, $a=1.0$ and a power level of β. The total sample size is N, and $T^{\alpha}_{k,\nu,\rho}$ is the required critical point.

	β	n_0	n	N	$T^{\alpha}_{k,\nu,\rho}$
$k=2$	0.95	31	23	77	1.96
	0.90	24	19	62	1.96
	0.80	19	14	47	1.98
	0.70	15	11	37	1.99
	0.60	12	9	30	2.01
$k=3$	0.95	38	23	107	2.11
	0.90	31	19	88	2.12
	0.80	25	14	67	2.13
	0.70	18	12	54	2.14
	0.60	17	9	44	2.16
$k=4$	0.95	44	23	136	2.22
	0.90	36	19	112	2.23
	0.80	26	15	86	2.23
	0.70	22	12	70	2.24
	0.60	17	10	57	2.25
$k=5$	0.95	49	23	164	2.30
	0.90	41	19	136	2.31
	0.80	30	15	105	2.31
	0.70	25	12	85	2.32
	0.60	20	10	70	2.33
$k=6$	0.95	55	23	193	2.37
	0.90	46	19	160	2.37
	0.80	35	15	125	2.38
	0.70	30	12	102	2.39
	0.60	24	10	84	2.40

Table 5.2

Table of the power levels at the least favourable configuration for the two-sided test procedure with known variance, k=3, a=1.0, and α=0.05. For fixed total sample sizes N=120 and N=76, the control sample size n_0 and the treatment sample size n are varied. $|T|^{\alpha}_{k,\infty,\rho}$ is the critical point, and $\hat{\lambda}$ is the value of λ at the least favourable configuration.

N	n_0	n	$\lvert T\rvert^{\alpha}_{k,\infty,\rho}$	$\hat{\lambda}$	n_0/n	power
76	37	13	2.378	0.11	2.846	0.772
76	34	14	2.376	0.13	2.429	0.786
76	31	15	2.373	0.14	2.067	0.796
76	28	16	2.369	0.16	1.750	0.799
76	25	17	2.364	0.18	1.471	0.798
76	22	18	2.357	0.20	1.222	0.789
76	19	19	2.349	0.22	1.000	0.772
76	16	20	2.338	0.24	0.800	0.744
76	13	21	2.323	0.26	0.619	0.699
120	54	22	2.376	0.15	2.455	0.944
120	51	23	2.374	0.16	2.217	0.947
120	48	24	2.372	0.17	2.000	0.950
120	45	25	2.369	0.18	1.800	0.951
120	42	26	2.367	0.20	1.615	0.951
120	39	27	2.363	0.21	1.444	0.950
120	36	28	2.359	0.23	1.286	0.947
120	33	29	2.355	0.24	1.138	0.943
120	30	30	2.349	0.25	1.000	0.937

Table 5.3

Table of the required sample sizes n_0 from the control and n from the treatments for the two-sided test procedure with $\nu=\infty$, k treatments, $\alpha=0.05$, $a=1.0$ and a power level of β. The total sample size is N, $|T|^{\alpha}_{k,\infty,\rho}$ is the required critical point, and $\hat{\lambda}$ is the value of λ at the least favourable configuration.

	β	n_0	n	N	$\lvert T\rvert^{\alpha}_{k,\infty,\rho}$	$\hat{\lambda}$
k=2	0.95	35	26	87	2.220	0.23
	0.90	29	21	71	2.220	0.21
	0.80	22	16	54	2.220	0.19
	0.70	17	13	43	2.219	0.18
	0.60	13	11	35	2.217	0.18
k=3	0.95	45	25	120	2.369	0.18
	0.90	36	21	99	2.368	0.18
	0.80	28	16	76	2.387	0.16
	0.70	23	13	62	2.390	0.15
	0.60	17	11	50	2.365	0.15
k=4	0.95	52	25	152	2.472	0.16
	0.90	42	21	126	2.471	0.16
	0.80	34	16	98	2.473	0.14
	0.70	28	13	80	2.473	0.12
	0.60	21	11	65	2.470	0.12
k=5	0.95	58	25	183	2.550	0.15
	0.90	48	21	153	2.550	0.14
	0.80	39	16	119	2.551	0.12
	0.70	32	13	97	2.552	0.12
	0.60	26	11	81	2.551	0.11
k=6	0.95	64	25	214	2.613	0.14
	0.90	52	21	178	2.612	0.13
	0.80	38	17	140	2.609	0.13
	0.70	31	14	115	2.609	0.11
	0.60	30	11	96	2.615	0.10

Table 5.4

A comparison between the lower bounds on the least favourable power with $\nu = 20$(top) and the least favourable powers with $\nu = \infty$(bottom), under square-root sample allocation, for different values of $h = \sqrt{N/(k+1)}\ a$. α is the size of the test, while k is the number of treatments.

α	h	k=2	3	4	5	6	7	8
0.10	2.0	0.319	0.286	0.266	0.252	0.242	0.233	0.226
		0.341	0.310	0.291	0.278	0.268	0.260	0.253
	2.5	0.438	0.397	0.372	0.354	0.340	0.329	0.320
		0.470	0.433	0.410	0.394	0.382	0.373	0.365
	3.0	0.567	0.524	0.496	0.476	0.461	0.449	0.438
		0.605	0.568	0.546	0.530	0.518	0.508	0.500
	3.5	0.692	0.651	0.625	0.606	0.592	0.580	0.570
		0.731	0.699	0.680	0.667	0.657	0.649	0.642
	4.0	0.798	0.765	0.743	0.728	0.716	0.706	0.698
		0.833	0.810	0.796	0.787	0.780	0.774	0.770
	4.5	0.879	0.855	0.840	0.828	0.820	0.813	0.807
		0.907	0.892	0.883	0.878	0.874	0.871	0.868
	5.0	0.934	0.919	0.909	0.902	0.896	0.892	0.888
		0.953	0.945	0.940	0.937	0.935	0.934	0.933
	5.5	0.967	0.959	0.953	0.949	0.946	0.945	0.944
		0.979	0.975	0.973	0.971	0.971	0.970	0.970
0.05	2.0	0.209	0.182	0.167	0.156	0.148	0.142	0.137
		0.232	0.206	0.191	0.181	0.173	0.167	0.162
	2.5	0.310	0.274	0.253	0.238	0.227	0.218	0.211
		0.347	0.314	0.295	0.281	0.271	0.263	0.257
	3.0	0.432	0.390	0.365	0.347	0.334	0.323	0.314
		0.482	0.446	0.425	0.410	0.400	0.391	0.384
	3.5	0.562	0.520	0.493	0.475	0.461	0.450	0.440
		0.621	0.587	0.568	0.554	0.545	0.537	0.531
	4.0	0.687	0.648	0.624	0.608	0.595	0.584	0.576
		0.746	0.719	0.703	0.693	0.686	0.681	0.677
	4.5	0.794	0.763	0.743	0.730	0.720	0.711	0.704
		0.845	0.827	0.816	0.810	0.806	0.802	0.800
	5.0	0.875	0.853	0.839	0.830	0.823	0.817	0.812
		0.915	0.904	0.898	0.895	0.893	0.891	0.890
	5.5	0.931	0.917	0.908	0.902	0.898	0.894	0.892
		0.958	0.952	0.949	0.948	0.947	0.947	0.947

Table 5.5

The exact size level of the test of H_0 at the common proportion $\bar{\pi}$, with k treatment populations, nominal size $\alpha = 0.10$, and the sample sizes N, n_0, n. $|T| \equiv |T|^{0.1}_{k,\infty,\frac{1}{1+\sqrt{k}}}$ is the critical point.

	N (n_0, n)	$\bar{\pi}=0.1$	$\bar{\pi}=0.2$	$\bar{\pi}=0.3$	$\bar{\pi}=0.4$	$\bar{\pi}=0.5$
$k = 2$	364 (150, 107)	0.107	0.101	0.101	0.102	0.102
$\lvert T\rvert = 1.927$	271 (113, 79)	0.100	0.099	0.100	0.100	0.100
	211 (87, 62)	0.101	0.099	0.102	0.101	0.098
	164 (68, 48)	0.096	0.095	0.098	0.101	0.102
	126 (52, 37)	0.099	0.087	0.101	0.101	0.099
$k = 3$	515 (188, 109)	0.098	0.099	0.101	0.100	0.102
$\lvert T\rvert = 2.087$	385 (142, 81)	0.097	0.102	0.098	0.100	0.100
	304 (112, 64)	0.098	0.102	0.097	0.098	0.099
	239 (89, 50)	0.095	0.103	0.101	0.106	0.099
	189 (69, 40)	0.112	0.100	0.102	0.102	0.103
$k = 4$	665 (221, 111)	0.100	0.098	0.101	0.099	0.106
$\lvert T\rvert = 2.198$	498 (166, 83)	0.098	0.103	0.102	0.100	0.100
	390 (130, 65)	0.108	0.103	0.101	0.100	0.104
	316 (104, 53)	0.102	0.105	0.101	0.096	0.100
	252 (84, 42)	0.100	0.101	0.104	0.098	0.097
$k = 5$	801 (246, 111)	0.102	0.102	0.101	0.100	0.099
$\lvert T\rvert = 2.282$	611 (191, 84)	0.101	0.103	0.100	0.101	0.101
	490 (150, 68)	0.104	0.099	0.103	0.100	0.101
	390 (120, 54)	0.102	0.108	0.104	0.102	0.101
	314 (99, 43)	0.107	0.098	0.101	0.100	0.100
$k = 6$	945 (273, 112)	0.100	0.103	0.101	0.100	0.100
$\lvert T\rvert = 2.350$	722 (212, 85)	0.103	0.100	0.102	0.101	0.101
	579 (165, 69)	0.106	0.104	0.104	0.101	0.104
	471 (135, 56)	0.107	0.103	0.103	0.100	0.102
	376 (112, 44)	0.105	0.102	0.100	0.100	0.101

Table 5.6

The exact value of β^* defined in equation (5.29) of the test with k treatment populations, nominal size $\alpha = 0.10$, the sample sizes N, n_0, n, and $a = 0.1$, 0.2. $|T| = |T|^{0.1}_{k,\infty,\frac{1}{1+\sqrt{k}}}$ is the critical point.

	N	n_0	n	a=0.1	a=0.2
$k = 2$	364	150	107	0.40	0.90
$\|T\| = 1.927$	271	113	79	0.32	0.80
	211	87	62	0.27	0.70
	164	68	48	0.23	0.60
	126	52	37	0.20	0.50
$k = 3$	515	188	109	0.38	0.90
$\|T\| = 2.087$	385	142	81	0.30	0.80
	304	112	64	0.26	0.70
	239	89	50	0.22	0.60
	189	69	40	0.20	0.50
$k = 4$	665	221	111	0.37	0.90
$\|T\| = 2.198$	498	166	83	0.29	0.80
	390	130	65	0.25	0.70
	316	104	53	0.21	0.60
	252	84	42	0.19	0.50
$k = 5$	801	246	111	0.35	0.90
$\|T\| = 2.282$	611	191	84	0.28	0.80
	490	150	68	0.24	0.70
	390	120	54	0.21	0.60
	314	99	43	0.18	0.50
$k = 6$	945	273	112	0.34	0.90
$\|T\| = 2.350$	722	212	85	0.28	0.80
	579	165	69	0.24	0.70
	471	135	56	0.21	0.60
	376	112	44	0.18	0.50

Chapter 6

Summary of results and directions for future research

6.1 Summary of results

The concept which runs through this thesis is that of the alternative hypothesis partitioning approach presented in chapter 1. Generally speaking, this idea is applicable to any test problem where power consideration is involved.

For the one-way fixed effects analysis of variance model, we propose the difference zones of the forms $\Theta^1 = \{ \mu : \max_{1 \le i \le k} |\mu_i - \bar{\mu}| \ge b_0 \}$ and $\Theta^2 = \{ \mu : \max_{1 \le i,j \le k} |\mu_i - \mu_j| \ge b_0 \}$, which are more intuitive than the traditionally used $\Theta^3 = \{ \mu : \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2 \ge b_0 \}$. For guaranteeing large power levels for Θ^i $(i = 1, 2)$, range type tests are usually superior to F-test. But for guaranteeing small power levels (close to size α) for Θ^i $(i = 1, 2)$, the F-test is usually superior to range type tests. The implementation of the Studentised range test and F-test for guaranteeing the power requirements for Θ^i is discussed fully in chapter 2. The problem of testing the equality of k Bernoulli success probabilities is discussed in chapter 4. A test is proposed, whose exact size and the power at the asymptotic least favourable configuration for $\{ \pi : \max_{1 \le i,j \le k} |\pi_i - \pi_j| \ge \delta \}$ can easily be computed.

In chapter 5, we consider the problem of comparing k treatments with a control. We show how to chose proper sample sizes from the treatments and the control to guarantee the power levels for the difference zones $\Theta_{A1} = \{ \mu : \max_{1 \le i \le k} (\mu_i - \mu_0) \ge a\sigma \}$ and $\Theta_{A2} = \{ \mu : \max_{1 \le i \le k} |\mu_i - \mu_0| \ge a\sigma \}$ by using Dunnett's one-sided and two-sided tests respectively. The similar problem for Bernoulli data is also consider in chapter 5.

In the next two sections, we discuss some possible directions for further research.

6.2 Some further research topics

In the preceding chapters we have discussed the problems of simultaneously testing the null hypothesis $H_0 : \mu_1 = \dots = \mu_k$, the equality of k treatment means, and the null hypothesis $H_0 : \mu_1 = \dots = \mu_k = \mu_0$, the equality of k treatment means and a control mean. The most obvious direction of further research is to consider testing the null hypothesis

$$H_0 : (\mu_1 , \dots , \mu_k) = (\mu_1^0 , , \dots , \mu_k^0)$$

against the general alternative hypothesis "not H_0", where $(\mu_1^0 , \dots , \mu_k^0)$ is a known constant vector. The usual test procedure for this problem is the Studentised maximum modulus test procedure introduced by Tukey (see Miller (1966)), which

$$\text{reject } H_0 \qquad \text{iff} \qquad \max_{1 \le i \le k} | \bar{X}_i - \mu_i^0 | / S > |m|_{k,\nu}^{\alpha} ,$$

where $\bar{X}_i$, $1 \le i \le k$ are assumed to be independently normally distributed random variables with means $\mu_1 , \dots , \mu_k$ and common variance σ^2, S^2 is a χ^2 estimator of σ^2 with ν degrees of freedom independent of $\bar{X}_1 , \dots , \bar{X}_k$, and $|m|_{k,\nu}^{\alpha}$ is the upper α-point of the Studentised maximum modulus distribution with parameters k, ν. The questions that need to be answered are then: first, how to guarantee the power level of the Studentised maximum modulus test for the difference zone of the form

$$\Theta_A = \{ \mu = (\mu_1 , \dots , \mu_k \} : \bigcup_{i=1}^{k} (|\mu_i - \mu_i^0| \ge a_i^0) \} ,$$

where a_i^0, $1 \le i \le k$, are preassigned positive constants; secondly, what is the form of the maximin test procedure for the above test problem, and the relationship between the maximin test and the maximum modulus test.

A second direction of further research could be to consider distributional assumptions other than normal or binomial. For example, the simultaneous comparison of k poisson populations, or simultaneous comparison of the location parameters of k uniform distributions.

A third direction we could take is to consider models with certain structures. For example, consider the simplest linear regression model

$$y_i = a + b\,x_i + e_i \;, \qquad 1 \le i \le k$$

where $\mathbf{e} = (\, e_1 \,, \,\ldots\, , \, e_k \,) \sim N_k(\mathbf{0}, \, \sigma^2\mathbf{I})$, and we are interested in testing the null hypothesis

$$\mathrm{H}_0 \;:\; (\, a, \, b \,) = (\, a_0 \,, \, b_0 \,)$$

against the general alternative "not H_0", where a_0, b_0 are constants. The problems are then: what is the proper form of the difference zone? which test procedure should be used for the above test problem? how may the power level for the employed test procedure and the defined difference zone be guaranteed? and how to design the experiment in the sense of choosing proper values of $x_1 \,, \,\ldots\, , \, x_k$ (in some interval) to maximise the least favourable power?

A fourth possible direction is to extend our ideas to the order restrictions model where the treatment means are known to have a certain ordering. Various test procedures have been developed for this problem, but little information is available on the power properties of such procedures.

Generally speaking, our ideas are applicable to any test problem where a power consideration is involved. An alternative approach to the multiple comparison problem is to design experiments to control confidence interval lengths; in this case, a sample size should be chosen so that the confidence intervals will simultaneously cover the true parameters and be sufficiently narrow with the required probability. This method has certain advantages since the probability of correct multiple comparisons inferences will be guaranteed. Some work on this topic has already been performed (see, for example, Hsu (1988)). In the next section we consider some non-parametric problems.

6.3 Non-parametric problems

Consider the problem of performing a hypothesis test on the location parameter of an unknown univariate distribution. The usual test procedure for this problem is the sign test which will guarantee type one error probability without any assumptions regarding the form of the unknown distribution. However, without such assumptions being made, very little can be guaranteed about the power function of the test procedure. One way to avoid this problem is to define a class of distributions which is believed to contain the unknown distribution in question, and then to obtain lower bounds on the power function for this class of distributions. The class of distributions should be wide enough to contain any distribution which may occur in practice, yet restrictive enough to produce useful bounds on the power function. The distribution in the class which achieves the lower bound on the power functions is called the **least favourable distribution.**

We consider here only the class of distributions with symmetric, unimodal densities. Unimodal densities have received considerable attention in statistical literature (see, for example, Dharmadhikari and Joag-dev (1988)) and have intuitive appeal, while the requirement of symmetry will in many cases be automatic when, for example, the data arises as the differences of i.i.d. random variables. This class also contains the normal distribution and other common distributions. For non-parametric test procedures, usually, only the size of the test is considered. In the following, we will shown how to assess the power properties of a test procedure by assuming that the real distribution is symmetric and unimodal.

Suppose we have data $X_1 , \dots , X_n$ which are i.i.d. random variables with an unknown mean μ, and we want to carry out a hypothesis test on the value of μ when the distribution of the X_i is only assumed to be unimodal and symmetric about μ. Below, we will discuss two approaches to this problem, namely, the sign test and the test based on the sample mean $\overline{X}$. We will consider only one-sided tests, as two-sided situations are essentially no different.

A size α one-sided sign test of

$$H_0 : \mu = 0 \quad \text{against} \quad H_A : \mu > 0$$

is to reject the null hypothesis if the number of positive X_i exceeds the critical value c, where c is chosen such that

$$P (B(n, \tfrac{1}{2}) \geq c) = \alpha ,$$

where $B(n, ½)$ is a binomial random variable with parameters n and ½. This test procedure is, in fact, valid for any distribution with median μ. However, the conditions of unimodality and symmetry allow a lower bound to be obtained for the power function. The power of the test is $P(\, B(n, p) \geq c\,)$ where

$$p = P(\, X_1 \geq 0) = 1 - ½P(\, |X_1 - \mu| \geq \mu\,)\,.$$

Notice that $P(\, B(n, p) \geq c\,)$ will increase as p increases, so in order to obtain a lower bound on the power function we only need to obtain the upper bound on

$$P\{\, |X_1 - \mu| \geq \mu\,\}$$

assuming the unimodality and symmetry of the distribution of X_1 about μ. Now we have the Gauss inequalities:

$$P(\, |Y| \geq d\,) \leq \begin{cases} 1 - \dfrac{d}{\sqrt{3}\sigma}\,, & d^2 \leq \dfrac{4}{3}\sigma^2 \\[2ex] \dfrac{4\sigma^2}{9d^2}\,, & d^2 \geq \dfrac{4}{3}\sigma^2 \end{cases}$$

where the random variable Y is assumed to have a distribution symmetric and unimodal about zero, σ^2 is the variance of Y, and d is any positive constant. The extremal distributions of the above inequalities are

$$d^2 \leq \frac{4}{3}\sigma^2 \quad \Rightarrow \quad \text{rectangular on } [-\sqrt{3}\sigma,\ \sqrt{3}\sigma]$$

$$d^2 \geq \frac{4}{3}\sigma^2 \quad \Rightarrow \quad \text{rectangular on } [-\frac{3}{2}d,\ \frac{3}{2}d] \text{ plus mass at } \mu$$

(See Karlin & Studden (1966) p.483). By applying these inequalities, a lower bound on the power function of the test procedure can be obtained in terms of μ/σ, where σ^2 is the variance of X_1. This lower bound is strict in the sense that there exists a unimodal symmetric distribution (rectangular with possibly a mass at the mode) for which the lower bound is attained.

We now consider the one-sided hypothesis test of

$$H_0 : \mu = 0 \qquad \text{against} \qquad H_A : \mu > 0$$

which rejects the null hypothesis whenever the sample mean $\bar{X} \geq c$. It is well known that the convolution of two symmetric unimodal densities is symmetric and unimodal (see, for example, Eaton (1987) p.7), and so the distribution of $\bar{X}$ is unimodal and symmetric about μ. Therefore, since the variance of $\bar{X}$ is σ^2/n, the Gauss inequalities allow the calculation of the following upper bounds on the size of the test procedure in terms of σ^2:

$$\text{size} = \tfrac{1}{2} P(\, |\bar{X}-\mu| \geq c \,) \leq \begin{cases} \dfrac{1}{2} - \dfrac{c\sqrt{n}}{2\sqrt{3}\sigma}\,, & c \leq \dfrac{2\sigma}{\sqrt{3n}} \\[2ex] \dfrac{2\sigma^2}{9nc^2}\,, & c \geq \dfrac{2\sigma}{\sqrt{3n}}\,. \end{cases}$$

These inequalities enable the critical point c to be chosen to guarantee a required size α; e.g. if $\alpha \leq 1/6$ (so that the bottom inequality is pertinent) then a critical value $c = \sigma\sqrt{2/(9n\alpha)}$ is required. Notice that this requires σ^2, or at least an upper bound for σ^2, to be specified.

The Gauss inequalities may also be applied to give a lower bound on the power function for $\mu \geq c$ (for $\mu < c$, there is no positive bound on the power function):

$$\begin{aligned} \text{power} &= P(\, \bar{X}-\mu \geq c-\mu \,) \\ &= 1 - \tfrac{1}{2} P(\, |\bar{X}-\mu| \geq \mu-c \,) \\ &\geq \begin{cases} \dfrac{1}{2} + \dfrac{(\mu-c)\sqrt{n}}{2\sqrt{3}\sigma}\,, & c \leq \mu \leq c + \dfrac{2\sigma}{\sqrt{3n}} \\[2ex] 1 - \dfrac{2\sigma^2}{9n(\mu-c)^2}\,, & \mu \geq c + \dfrac{2\sigma}{\sqrt{3n}}\,. \end{cases} \end{aligned}$$

If the critical value given above for $\alpha \leq 1/6$ is used, then the power function $\beta(\delta)$, where $\delta = \mu/\sigma$, has the following lower bounds

$$\beta(\delta) \geq \begin{cases} \dfrac{1}{2} + \dfrac{3\delta\sqrt{n}-\sqrt{2/\alpha}}{6\sqrt{3}}\,, & \dfrac{1}{3}\sqrt{\dfrac{2}{n\alpha}} \leq \delta \leq \dfrac{2}{\sqrt{3n}} + \dfrac{1}{3}\sqrt{\dfrac{2}{n\alpha}} \\[2ex] 1 - \dfrac{2}{(3\delta\sqrt{n}-\sqrt{2/\alpha})^2}\,, & \delta \geq \dfrac{2}{\sqrt{3n}} + \dfrac{1}{3}\sqrt{\dfrac{2}{n\alpha}}\,. \end{cases}$$

Thus it can be seen that the assumptions of unimodality and symmetry are sufficient to enable the employment of this simple test procedure for which the critical points and power properties are readily calculable without the use of any tables. However, the problem arising here is that the lower bound on the power function is not traceable, i.e. $\bar{X}$ cannot have the least favourable distribution. This is because the rectangular distribution (with possibly a mass at the middle) cannot arise as the convolution of any unimodal symmetric distribution, which can be seen from the fact that as $n \to \infty$, the distribution of $\bar{X}$ tends to normality regardless of the distribution of the X_i. So, inequalities similar to the Gauss inequalities need to be derived for the distribution formed as the n-th convolution of a unimodal symmetric distribution. Such inequalities should be tighter than those derived from the Gauss inequalities. If such inequalities could be found, then the lower bound of the power function could be improved. It seems, however, that such inequalities are not simple to derive. Some related problems have been studied in the literature, see, for example, Hoeffding (1955), Hoeffding & Shrikhande (1955), and Rustagi (1957).

Above, we considered the problem for the class of distributions which are symmetric and unimodal. We can also consider the problem for other classes of distributions, for example, the class of unimodal distributions. Another problem is how to extend the above ideas to the multiple comparison problems. Suppose X_{ij}, $1 \le i \le k$, $1 \le j \le n$ are k independent samples with n independent observations in each samples. The observations are assumed to have the distributions

$$X_{ij} \sim F(\, x-\theta_i \,)\ ,\quad 1 \le i \le k,\ 1 \le j \le n,$$

where $F(.)$ is a fixed (unknown) distribution function. The parameters θ_i are not uniquely defined because of the arbitrariness of the distribution. However, the differences $\theta_i - \theta_j$ are uniquely defined, and these are the parametric quantities of interest. The null hypothesis is

$$\mathrm{H}_0 \;:\; \theta_1 = \theta_2 = \ldots = \theta_k\ .$$

The alternative is simply "not H_0". The usual non-parametric test procedures for this problem are based on the rank or signed rank (see Miller(1981) for details). We, however, want to use the following test procedure

$$\text{reject } \mathrm{H}_0 \quad \text{iff} \quad \max_{1 \le i,j \le k} |\bar{X}_i - \bar{X}_j| \ge c\ ,$$

where $\bar{X}_i$ is the sample mean of the ith population, and c is a critical point. To determine the size of this test, we need some inequality in the form

$$P\left\{ \max_{1 \le i,j \le k} |\bar{X}_i - \bar{X}_j| \ge c \right\} \le \ ? \tag{6.1}$$

We propose the above test procedure because of its intuitive appeal and the ease of deriving the simultaneous confidence intervals for the pairwise differences $\theta_i - \theta_j$ from the test procedure. The derivation of an inequality along the lines of

(6.1) is of interest by itself.

In Theorem 6.1 below, we give one such inequality, which is an immediate consequence of the following two lemmas.

Lemma 6.1 (Dharmadhikari & Joag-dev (1985))

Let Y be a random variable with a unimodal distribution. Let $a \in \mathbf{R}$, $r > 0$, and $\tau_r = E(|Y-a|^r)$. Then, for every $c > 0$,

$$P\left\{ |Y-a| \geq c \right\} \leq \max\left\{ \frac{s\tau_r - c^r}{(s-1)c^r},\ [\frac{r}{r+1}]^r \frac{\tau_r}{c^r} \right\},$$

where s satisfies the conditions

$$s > r+1 \quad \text{and} \quad s(s-r-1)^r = r^r .$$

Lemma 6.2 (David (1970) p.46-50)

Let $Z_1, \ldots, Z_k$ be i.i.d. random variables with distribution $G(.)$. If $G(.)$ is continuous and strictly increasing, and its variance equals one, then

$$E\left\{ \max_{1 \leq i,j \leq k} |Z_i - Z_j| \right\} \leq k\,B ,$$

where

$$B = \sqrt{\frac{2}{2k-1}[\,1 - 1/\binom{2k-2}{k-1}\,]} .$$

The extreme c.d.f. is determined by

$$x = \frac{G^{k-1} - (1-G)^{k-1}}{B}, \quad x \in (-\frac{1}{B}, \frac{1}{B}) .$$

Combination of the above two lemmas gives the following theorem, which enables us to obtain an upper bound for the probability

$$P\left\{ \max_{1 \leq i,j \leq k} |\bar{X}_i - \bar{X}_j| \geq c \right\}.$$

Theorem 6.1 Suppose X_{ij} ($1 \leq i \leq k,\ 1 \leq j \leq n$) $\overset{i.i.d.}{\sim}$ $F(x-\theta)$. If

(i) $F(.)$ is continuous, strictly increasing, and its variance exists and is equal to σ^2.

(ii) $R_k \equiv \max\limits_{1 \le i,j \le k} |\bar{X}_i - \bar{X}_j|$ is unimodally distributed.

then

$$P\{\ R_k \ge c > 0\ \} \le \max \left\{ \frac{(1+\sqrt{2})A-c}{\sqrt{2}c},\ \frac{A}{2c} \right\},$$

where $A = \sigma k B/\sqrt{n}$, and B is as defined in Lemma 6.2.

Proof of Theorem 6.1

By taking $r = 1$, $a = 0$, and $Y = R_k$ in Lemma 6.1, we have

$$P\{\ R_k \ge c\ \} \le \max \left\{ \frac{(1+\sqrt{2})ER_k-c}{\sqrt{2}c},\ \frac{ER_k}{2c} \right\}.$$

This, and Lemma 6.2 with $Z_i = \sqrt{n}\bar{X}_i/\sigma$ immediately give the results of the theorem #

However, the problem here is that the upper bound for the probability $P\{\ R_k \ge c\ \}$ given in the theorem is not strict. This is because the extreme distribution for Lemma 6.2 is not the extreme distribution for Lemma 6.1. So the inequality in the theorem could be improved. Another problem in deriving such inequalities is whether conditions such as unimodality should be imposed on R_k or $G(.)$. If conditions are imposed on R_k, then what kind of $G(.)$ will form such R_k?

The above problem arises from the "non-parametricisation" of the Studentised range test. Many similar problems can also arise. Further research could be to find the lower bound on the power functions of such test procedures.

Appendix

This appendix contains some definitions and theorems which are required in the preceding chapters.

Definition 1 (majorisation)

Let $\mathbf{a} = (a_1 , a_2 , \dots , a_k)$ and $\mathbf{b} = (b_1 , b_2 , \dots , b_k)$. In addition let $a_{[1]} \le a_{[2]} \le \dots \le a_{[k]}$ and $b_{[1]} \le b_{[2]} \le \dots \le b_{[k]}$ denote the ordered components of $\mathbf{a}$ and $\mathbf{b}$. Then $\mathbf{a}$ is said to majorise $\mathbf{b}$ (written as $\mathbf{a} \gg \mathbf{b}$), if

$$\sum_{i=1}^{k} a_i = \sum_{i=1}^{k} b_i ,$$

$$\sum_{i=r}^{k} a_{[i]} \ge \sum_{i=r}^{k} b_{[i]} , \qquad 2 \le r \le k .$$

Definition 2 (Schur-concavity)

The function $f(\mathbf{x}) : \mathbf{R}^k \to \mathbf{R}$ is called Schur-concave if

$$\mathbf{a} \gg \mathbf{b} \quad \Rightarrow \quad f(\mathbf{a}) \le f(\mathbf{b})$$

for all $\mathbf{a} , \mathbf{b} \in \mathbf{R}^k$.

Some sufficient or necessary and sufficient conditions for a function to be Schur-concave can be seen in Marshall & Olkin (1979).

Theorem A1 (Marshall & Olkin (1974))

Suppose that $\mathbf{x} = (x_1 , \dots , x_k)$ has a Schur-concave joint density function $f(\mathbf{x})$. Let $\mathbf{A} \subset \mathbf{R}^k$ be a Lebesgue-measurable set such that

$$\mathbf{a} \in \mathbf{A} , \ \mathbf{a} \gg \mathbf{b} \quad \Rightarrow \quad \mathbf{b} \in \mathbf{A} \tag{A1}$$

Then

$$\int_{\mathbf{A}} f(\mathbf{x} - \boldsymbol{\theta}) \, d\mathbf{x}$$

is a Schur-concave function of $\boldsymbol{\theta}$.

Note A1 Condition (A1) in Theorem A1 is satisfied whenever **A** is a permutation invariant convex set (see Marshall & Olkin (1974)).

Definition 3 (log-concavity)

The function $f(\mathbf{x}) : \mathbf{R}^k \to [0, \infty)$ is called log-concave if

$$f(\,\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\,) \geq [f(\,\mathbf{x}\,)]^{\alpha}\,[f(\,\mathbf{y}\,)]^{1-\alpha}$$

for all $0 \leq \alpha \leq 1$, and $\mathbf{x},\ \mathbf{y} \in \mathbf{R}^k$.

Theorem A2 (Prekopa (1973))

Let f be defined on $\mathbf{R}^m \times \mathbf{R}^n$ and suppose that f is log-concave. Then, the function h defined on $\mathbf{R}^m$ by

$$h(\mathbf{u}) = \int_{\mathbf{R}^n} f(\mathbf{u},\, \mathbf{v})\; d\mathbf{v}\ ,$$

and assumed to be finite for $\mathbf{u} \in \mathbf{R}^m$, is log-concave.

Definition 4 (unimodality)

The function $f(\mathbf{x}) : \mathbf{R}^k \to [0, \infty)$ is called unimodal if the set

$$D_u = \{\ \mathbf{x} \mid f(\mathbf{x}) \geq u\ \} \subset \mathbf{R}^k$$

is convex for all $u \geq 0$.

Theorem A3 (Mudholkar (1969))

If the function $f_0(x) : \mathbf{R} \to [0, \infty)$ is log-concave, then the function defined by

$$f(\mathbf{x}) = \prod_{i=1}^{k} f_0(x_i) :\ \mathbf{R}^k \to [0, \infty)$$

is a unimodal function.

Theorem A4 (Anderson (1955))

Let the function $f(\mathbf{x}) : \mathbf{R}^k \to [0, \infty)$ be symmetric about the origin and unimodal. Let $A \subset \mathbf{R}^k$ be symmetric about the origin and convex. If $\int_A f(\,\mathbf{x} + \lambda\mathbf{y}\,)\; d\mathbf{x} < \infty$ (in the Lebesgue sense), then

$$\int_A f(\mathbf{x} + \lambda\mathbf{y})\, d\mathbf{x} \geq \int_A f(\mathbf{x} + \mathbf{y})\, d\mathbf{x}$$

for all $\mathbf{y} \in \mathbf{R}^k$ and all $\lambda \in [0, 1]$.

Theorem A5 (Hardy, Littlewood & Polya (1952) p.261)

If $a_i, b_i,\ i = 1, \dots, k$ are two sets of numbers, and $a_{[1]} \leq \dots \leq a_{[k]}$, $b_{[1]} \leq \dots \leq b_{[k]}$ denote the ordered $a_i, b_i,\ 1 \leq i \leq k$, then

$$\sum_{i=1}^{k} a_{[i]}\, b_{[k-i+1]} \leq \sum_{i=1}^{k} a_i\, b_i \leq \sum_{i=1}^{k} a_{[i]}\, b_{[i]} \ .$$

References

Anscombe, F.J. (1948), "The transformation of poisson, binomial and negative-binomial data", *Biometrika,* 35, 246-254.

Bechhofer, R.E. (1954), "A single-sample multiple decision procedure for ranking means of normal populations with known variances", *The Annals of Mathematical Statistics,* 25, 16-39.

Bechhofer, R.E. (1969), "Optimal allocation of observations when comparing several treatments with a control", in Multivariate Analysis-II (ed. P.R. Krishnaiah), New York: Academic Press, 463-473.

Bechhofer, R.E., and Dunnett, C.W. (1988), "Tables of percentage points of multivariate Student t distributions", in Selected Tables in Mathematical Statistics.

Bechhofer, R.E., and Nocturne, D.J.M. (1972), "Optimal allocation of observations when comparing several treatments with a control, II: two-sided comparisons", *Technometrics,* 14, 423-436.

Bechhofer, R.E., and Tamhane, A.C. (1983), "Design of experiments for comparing treatments with a control: tables of optimal allocations of observations", *Technometrics,* 25, 87-95.

Bhapkar, V.P., and Somes, G.W. (1976), "Multiple comparisons of matched proportions", *Communications in Statistics,* A5, 17-25.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975), "Discrete Multivariate Analysis: Theory and Practice", MIT Press.

Bristol, D.R. (1989), "Designing clinical trials for two-sided multiple comparisons with a control", *Controlled Clinical Trials,* 10, 142-152.

Dasgupta, P. (1968), "Tables of the non-centrality parameter of F-test as a function of power", *Sankhya,* B-30, 73-82.

David, H.A. (1970), "Order Statistics", Wiley & Sons.

David, H.A., Lachenbruch, P.A., and Brandis, H. P. (1972), "The power function of range and Studentised range tests in normal samples", *Biometrika,* 59, 161-168.

Dharmadhikari, S.W., and Joag-dev, K. (1985), "The Gauss-Tchebyshev inequality for unimodal distributions", *Theor. Probability Appl.,* 30, 867-871.

Dharmadhikari, S., and Joag-dev, K. (1988), "Unimodality, Convexity, and Applications", Academic Press.

Dunnett, C.W. (1955), "A multiple comparison procedure for comparing several treatments with a control", *Journal of the American Statistical Association,* 50, 1096-1121.

Dunnett, C.W. (1964), "New tables for multiple comparisons with a control", *Biometrics,* 20, 482-491.

Eaton, M.L. (1987), "Lectures on Topics in Probability Inequalities", CWI Tract 35, Amsterdam, Netherlands.

Edwards, D., and Berry, J.J. (1987), "The efficiency of simulation-based multiple comparisons", *Biometrics,* 43, 913-928.

Fox, M. (1956), "Charts of the power of the F-test", *The Annals of Mathematical Statistics,* 27, 484-497.

Goodman, L.A. (1964), "Simultaneous confidence intervals for contrasts among multinomial populations", *The Annals of Mathematical Statistics,* 35, 716-725.

Gupta, S.S., and Panchapakesan, S. (1979), "Multiple Decision Procedures", Wiley & Sons.

Halperin, M., Greenhouse, S.W., Cornfield, J., and Zalokar, J. (1955), "Tables of percentage points for the Studentised maximum absolute deviate in normal samples", *J. Amer. Statist. Assoc.,* 50, 185-195.

Hardy, G.H., Littlewood, J.E., and Polya, G. (1952), "Inequalities", Second Edition, Cambridge University Press.

Hawkins, D.M., and Perold, A.F. (1977), "On the joint distribution of left- and right-sided outlier statistics", *Utilitas Mathematica,* 12, 129-143.

Hayter, A.J., and Liu, W. (1989), "Multiple comparisons with a control: sample size determination", Manuscript, School of Mathematical Sciences, University of Bath.

Hayter, A.J., and Liu, W. (1989), "Optimal test procedures for comparing several location parameters", submitted for publication.

Hayter, A.J. and Liu, W. (1990a), "The power function of the Studentised Range test", *The Annals of Statistics,* to appear.

Hayter, A.J., and Liu, W. (1990b), "Power assessment for tests of equality of several proportions", *Communications in Statistics,* to appear.

Healy, W.C., Jr. (1956), "Two-sample procedures in simultaneous estimation", *The Annals of Mathematical Statistics,* 27, 687-702.

Hochberg, Y., and Lachenbruch, P.A. (1976), "Two stage multiple comparison procedures based on the Studentised range", *Communications in Statistics,* A5(15), 1447-1453.

Hochberg, Y., and Tamhane, A.C. (1987), "Multiple Comparisons Procedures", Wiley, New York.

Hoeffding, W. (1955), "The extrema of the expected value of a function of independent random variables", *The Annals of Mathematical Statistics,* 26, 268-275.

Hoeffding, W., and Shrikhande, S.S. (1955), "Bounds for the distribution function of a sum of independent identically distributed random variables", *The Annals of Mathematical Statistics,* 26, 439-449.

Hsu, J.C. (1989), "Sample size computation for designing multiple comparison experiments", *Computational Statistics & Data Analysis,* 7, 79-91.

Hurn, M.A. (1989), "A study of the power functions of some optimal simultaneous inference procedures by exact evalution and simulation techniques", M.Sc. thesis, School of Mathematical Sciences, University of Bath.

Karlin, S., and Studden, W. (1966), "Tchebycheff Systems: with Applications in Analysis and Statistics", Wiley & Sons.

Kastenbaum, M.A., Hoel, D.G., and Bowman, K.O. (1970), "Sample size requirements: one-way analysis of variance", *Biometrika,* 57, 421-430.

Knoke, J.D. (1976), "Multiple comparisons with dichotomous data", *Journal of the American Statistical Association,* 71, 849-853.

Kunert, J. (1988), "On the power of tests for multiple comparison of three normal means", Manuscript, Department of Mathematics, Fachbereich IV, Universitat Trier.

Marshall, A.W., and Olkin, I. (1974), "Majorisation in Multivariate Distributions", *The Annals of Statistics,* 2, 1189-1200.

Marshall, A.W., and Olkin, I. (1979), "Inequalities: Theory of Majorization and its Applications", New York: Academic Press.

Miller, R.G., Jr. (1981), "Simultaneous Statistical Inference", Second Edition, Springer-Verlag New York Inc.

Moshman, J. (1958), "A method for selecting the size of the initial sample in Stein's two sample procedure", *The Annals of Mathematical Statistics,* 29, 1271-

1275.

Mudholkar, G.S. (1969), "A generalized monotone character of density functions and moments of statistics from some well-known populations", *Ann. Inst. Statist. Math.,* 21, 277-285.

Pearson, E.S., and Hartley, H.O. (1951), "Charts of the power function for analysis of variance tests, derived from the non-central F-distribution", *Biometrika,* 38, 112-130.

Prekopa, A. (1973), "On logarithmic concave measures and functions", *Acta. Scient. Mat. (Szeged),* 34, 335-343.

Rustagi, J.S. (1957), "On minimizing and maximizing a certain integral with statistical applications", *The Annals of Mathematical Statistics,* 28, 309-328.

Scheffe, H. (1959), "The Analysis of Variance", New York, Wiley.

Seelbinder, B.M. (1953), "On Stein's two-stage sampling scheme", *The Annals of Mathematical Statistics,* 24, 640-647.

Stein, C. (1945), "A two-sample test for a linear hypothesis whose power is independent of the variance", *The Annals of Mathematical Statistics,* 16, 243-258.

Tiku, M.L. (1967), "Tables of the power of the F-test", *Journal of the American Statistical Association,* 62, 525-539.

Tong, Y.L. (1980), "Probability Inequalities in Multivariate Distributions", New York: Academic Press.

Wald, A. (1942), "On the power function of the analysis of variance test", *The Annals of Mathematical Statistics,* 13, 434-439.

Williams, D.A. (1988), "Tests for differences between several small proportions", *J.R. Statist. Soc. C,* 37, 421-434.